

More strings for British research

The British research enterprise has escaped yet another radical reorganization, that which would have been entailed in the creation of a single research council. But it may yet be sorry.

DEMORALIZED British researchers at academic and cognate institutions have been surprisingly indifferent to the scheme, advocated last year by a committee under Mr J.R.S. Morris, that the five existing research councils should be rolled together. It is as if, like punchdrunk prizefighters, they have become self-preoccupied with pain, but uncaring about its source. Yet the issue is important if the enterprise is ever to recover from the past two decades of mismanagement and deprivation. And now, in any case, the issue has been settled (see page 294). The British government has set its face against upheaval, but will instead expect a reconstituted (and smaller) Advisory Board for the Research Councils (ABRC) to intervene more directly than has been its habit in the affairs of the research councils.

The obvious benefit of this decision is that another major upheaval will be avoided. That is something genuinely to be grateful for. Particular research councils, understandably jealous of the remnants of their autonomy, will also be at first relieved. For a while, at least, they can hope that the future will be an extrapolation of the past. And is it not true that the occasion of the Morris recommendations, principally the turf battle between the Medical Research Council and the Science and Engineering Research Council over the right to support non-medical biology and its applications, could be settled without the need for legislation, with all the hostages to fortune that would create? The British government similarly has reasons to welcome the quiet life, with the uncertain prospect of a general election just two years ahead.

But the new arrangement could prove to be less palatable than it now appears. Everything will hang on how ABRC sets about its larger task. In principle, the research council system has two functions — to support basic research in academic institutions and to carry out what is fashionably called strategic research of relevance in designated fields such as agriculture, medicine and the natural environment. Within living memory, in 1971, the then government (of the same stripe as the present) enthusiastically endorsed Lord Rothschild's notion that government departments should stand proxy for the ultimate taxpayer beneficiaries of strategic research, but the departments have consistently ducked their responsibilities, notably in agriculture. The research councils have not been able so lightly to shrug off their responsibilities. They have commitments to research institutes and their

staffs, while increasing proportions of what they spend are earmarked for particular fields of research.

The outstanding advantage of a unified research council is that it could have moulded its policies to a reappraisal of what is now the British national interest. Two objectives are conspicuous and generally agreed. First, Britain is heading for a shortage of technical skill that will be avoided only by the generous and ample support of graduate students and postdoctoral fellows. That requires that dignity should be restored to the research profession, whose penury has been too well advertised in the past decade. Second, the foundations of a healthy research community need to be rebuilt, in the hope that the British enterprise can recover some of its old sparkle. That requires that the research councils should spend a much larger proportion of their funds (nearly £900 million a year) on research projects conceived of by their clients, not themselves. There should be a third objective, less widely appreciated, that the shocking neglect of the social sciences in the past few years is made good.

Quite how even a streamlined ABRC can tackle these problems is far from clear. The council will remain advisory to the Department of Education and Science, not to the government as a whole. (The newest version of a central advisory body, the Advisory Council on Science and Technology, seems even less active than its predecessor, despite having the prime minister as chairman). It will have no executive authority, with the result that it will be able to persuade research councils to change their ways only by threatening that their annual budgets will be accompanied by explicit instructions from above, which will further demean the system as a whole while sanctifying the practice of earmarking research funds for particular projects.

The most serious danger in the new arrangements is that the link between the research community and its ultimate managers, one of the few sources of hope in a devastating decade, will be further attenuated. The other is that the whims of politicians, for projects as different as information technology and research in and around the Falklands Islands, will come to dominate British researchers' working lives. The new ABRC will need guile as well as courage to improve and not to worsen Britain's prospects in research. All that should matter now is that people's morale should be restored by a sense that their own good ideas can be prosecuted successfully. □



No forced merger for UK research councils

- Recommendation turned down
- Closer links likely for AFRC and NERC

London

The five UK research councils will not after all be merged into a single National Research Council, as was recommended in the Morris report, but a pared-down Advisory Board for the Research Councils (ABRC) will instead take a more interventionist role in disputes between them, according to an announcement made this week by John MacGregor, Secretary of State for Education and Science.

From 1 April, ABRC's membership will be cut from 26 to 14 people. The restructured body will be given "a more explicit remit to improve . . . joint working among the research councils", said MacGregor. Sir David Phillips will continue as chairman, and it is understood that the heads of the research councils will play a fuller role in the workings of the board. At present, some issues are discussed by the 'independent' members of ABRC before being referred to the full board. It is hoped that this change will mean that disputes between the research councils can be thrashed out more effectively, before ABRC submits advice to the government. The full membership of the revamped ABRC should be announced within the next few weeks.

The restructuring of ABRC is proposed as a compromise solution to the problems of coordinating the activities of the research councils and the overlaps of responsibility for some subject areas, which prompted the ABRC's Morris committee report, published in July 1989. The report's recommendation to merge the research councils had met vigorous opposition, notably from within the Medical Research Council (MRC).

The advice finally submitted to MacGregor followed from a second ABRC study, involving the heads of the research councils. Phillips, in a letter accompanying the ABRC's advice, regards the restructuring of ABRC as a good compromise solution to the problems raised in the Morris report, saying "this change would go a very long way towards overcoming the coordination difficulties . . . , whilst avoiding the disruption, transitional costs and need for legislation associated with more radical proposals".

The solution also seems acceptable to some opponents of the original Morris proposals. Max Perutz, of the MRC Laboratory of Molecular Biology at Cambridge, a vocal critic of plans to merge the research councils, says he is "very relieved that the government has

decided against a single National Research Council. This would have been a disaster."

But some reorganization of the research councils themselves is still a possibility. MacGregor is considering proposals to improve coordination between the Agricultural and Food Research Council

MILITARY RESEARCH

Cold War thaw strikes home

Boston

In an announcement widely seen as a harbinger of the changes coming for US military research, the Draper Laboratory in Cambridge, Massachusetts, last week said it will lay off 145 employees, including 75 scientists and engineers. The laboratory is a private, military research centre affiliated with the Massachusetts Institute of Technology until 1973 and relies on Defense Department contracts for 90 per cent of its budget.

Over the past seven months, the Draper Laboratory has trimmed its staff by about one fifth through layoffs or mandatory early retirement programmes. Kathleen Granchelli, a spokeswoman for the laboratory, said that some of the cutbacks had been expected in the light of the completion of much of the laboratory's work on the Trident II missile guidance system. But she added that programmes that the laboratory had hoped would supplant the Trident work have now been reduced, delayed or eliminated. And, she said, there is "no question that cuts in the defence budget will impact the lab in the future".

The cuts at Draper have been followed by an announcement of layoffs, losses and budget cuts at another Boston-based military contractor, M/A-Com Inc., a well-known defence electronics company that relies on the military for some 80 per cent of its business. M/A-Com announced an expected quarterly loss of \$115 million, and a layoff of more than ten per cent of its 6,500 employees as a result of "rapidly changing conditions in the defence electronics marketplace", says the company's chairman Thomas Vanderslice.

Robert Costello, a former Under Secretary of Defense for Acquisitions in the Reagan and Bush administrations and now a senior fellow at the Hudson Institute, a private research organization, says that virtually all military contractors are in the same situation. But he argues that it

(AFRC) and the Natural Environment Research Council (NERC). A merger between the two councils was first suggested in March 1988 by Sir Hugh Fish, then NERC chairman, and subsequently supported by the House of Lords Select Committee on Science and Technology.

But NERC has since retreated from support for a full merger, and now wishes to take over AFRC's responsibilities for environmental science. This particular option is not favoured by AFRC. Professor Bill Stewart, secretary of AFRC, says "AFRC has been consistent in the view that there would be merit in AFRC and NERC coming together".

Peter Aldhous

would be better if military research and development were protected from cuts.

Although many of the Pentagon's new weapons systems may need to be cut in the light of changing relationships with the Eastern bloc, he stresses that the military research and development establishment in the United States is a "tremendous and unique infrastructure" that has cost taxpayers "trillions" of dollars. "We had better figure out how to use it and redirect it appropriately before we destroy it."

One prospect, according to Costello, lies in 'dual-use' research which can provide both military and commercial benefits. His views are shared by many in the defence industry, including William Burnett, vice-president of the private consulting firm Charles River Associates, who specializes in advising defence contractors about weapons procurement policy. Burnett sees many military cutbacks as an "inevitable" consequence of the Defense Department's support for the development of so many expensive new weapons systems during the 1980s. "Essentially", he says, "there was no way we could continue to finance all of these projects through production at the same time." Now, he says, however, these trends have been compounded by the "hard times" besetting the Defense Department.

But not all observers are disheartened by a downturn in military spending. Columbia University professor Seymour Melman welcomes a reduction in military funding because he says the Defense Department has siphoned off talent and resources that can now be used more productively elsewhere. "The military is overwhelmed with a wave of panic" at the prospect of the "winding down of the cold war and the arms race", Melman says. The current emphasis on basic and 'dual-use' research he sees as merely "a desperate effort to try to diversify to protect their dominant position".

Seth Shulman

More riches for Japan

Tokyo

SPACE science, astronomy and research on the global environment receive strong backing from the Japanese government in an expansionary budget for science and technology released last week.

The Ministry of Education, Science and Culture (MESC), the Ministry of International Trade and Industry (MITI), the Environment Agency and the Ministry of Health and Welfare all receive significant increases for science-related items, as did the Science and Technology Agency (see *Nature* 343, 195; 18 January 1990). Several new projects are provided for.

The budget was approved by the Ministry of Finance and the cabinet at the end of last month and will run from April.

Even though some delay is now possible because of resistance from the newly strengthened opposition parties in the Diet, the outlays for science and technology are unlikely to be much affected.

Japan's astronomers seem set to achieve their dream of building a huge 7.5-metre infrared telescope on Hawaii after years of lobbying. The Ministry of Finance has approved a tiny budget of ¥13 million (\$90,000) in preparatory funds for the telescope, construction of which is eventually expected to cost about ¥45,000 million (\$320 million). But Makoto Kinoshita of MESC's science and international affairs bureau stresses that there are many hurdles still to be overcome before the project is realized.

More clear-cut is approval of ¥845 million (\$6 million) to build a heliograph at the Nobeyama Radio Observatory. The heliograph will consist of a T-shaped array of 76 80-cm parabolic antennas, and is expected to be completed in two years at a cost of ¥1,900 million.

MESC's budget for space science decreases slightly in 1990 but it contains an important new allocation of ¥1,850 million (\$13 million) to develop a new solid-fuel rocket for the Institute of Space and Astronautical Sciences (ISAS). The rocket will have three times the power of the institute's present biggest rocket, the MU-3SII, which this week is expected to send Japan's first probe to the Moon. The new rocket will take four years to develop at a cost of about ¥20,000 million.

The most striking feature of MITI's budget is a huge new outlay of ¥6,274 million (\$45 million) to develop new technology and products to protect the global environment. MITI plans to use the funds at a new institute that will be set up as a foundation with donations from industry later this year.

The Environment Agency also gets a huge boost in its budget for the global environment. Most of the extra funds go to new research grants (¥1,200 million)

for universities and research institutes in Japan and overseas, and to the establishment of a new centre for global environment research at the agency's National Institute for Environmental Studies in Tsukuba (¥285 million).

MESC also gets new funds to reorganize the Research Institute of Atmospheres at Nagoya University into the provisionally named Solar-Earth Environment Research Institute where fluxes of energy and materials from the Earth and Sun and their effect on the global environment will be studied. The National Institute of Polar Research, which until now has devoted its attention to the Antarctic, will establish an Arctic research centre. And the Sand Dune Research Institute of Tottori University will be reorganized to tackle research on the deserts of the world as well as its current research of local sand dunes.

MITI continues to pour money into the development of superconductors and into its fifth generation computer project which will end in 1992. The budget for a small unmanned space platform, the Space Flyer Unit, being jointly developed by MITI, ISAS and STA for launch in fiscal year 1994, also rises as it nears completion.

Funding for a ten-year project to study cancer initiated in 1984 with the backing of former prime minister Yasuhiro Naka-

One step forward, one step back

Washington

THE space shuttle Columbia landed at Edwards Air Force Base in southern California early last Saturday morning, a day late because of bad weather but with the Long Duration Exposure Facility (LDEF) safely stowed in its cargo bay. LDEF will be taken back to the Kennedy Space Center before it is removed from Columbia and its numerous onboard experiments unpacked and analysed (see *Nature* 342, 847; 1989).

The day before Columbia landed, the National Aeronautics and Space Administration announced that the launch of the Hubble Space Telescope has been set back once more, this time from 26 March to 19 April. Engineers decided to remove one of the solid rocket booster and replace one segment of it because their records did not show whether or not one of the O-ring joints had been fully tested for leaks.

David Lindley

sone (see *Nature* 310, 264; 1984) has been rising dramatically in recent years. The project is jointly funded by MESC, STA and the Ministry of Health and Welfare. But possibly the most significant item in the Ministry of Health and Welfare budget is a small outlay of about ¥300 million (\$2 million) to begin a project to sequence human oncogenes.

David Swinbanks

JAPANESE SCIENCE BUDGET

	1987	1988	1989	1990	(% change from 1989)
	(thousand million yen)				
<i>Ministry of Education, Science and Culture</i>					
Grants-in-aid of research	45.1	48.9	52.6	55.8	(+ 6.1%)
Government/industry research	7.9	8.9	10.3	11.5	(+ 11.9%)
Donations from industry	NA	NA	33.2	39.2	(+ 18.0%)
Domestic research fellowships	1.1	1.5	1.9	2.2	(+ 14.1%)
Nuclear fusion	7.7	7.6	8.5	8.9	(+ 4.1%)
Accelerator physics (TRISTAN)	12.9	16.5	15.9	16.3	(+ 2.3%)
Space science	11.8	19.8	20.8	18.0	(- 13.3%)
Astronomy	NA	NA	NA	0.9	—
Global environment	NA	NA	3.0	3.9	(+ 30.6%)
Earth science	2.1	2.1	2.2	2.3	(+ 4.6%)
Antarctic research	2.9	2.8	2.9	5.1	(+ 74.3%)
International exchange	5.5	6.6	5.8	6.2	(+ 6.1%)
<i>Environment Agency</i>					
Global environment	NA	0.3	0.9	2.1	(+142.0%)
<i>Ministry of International Trade and Industry (MITI)</i>					
Total R&D budget	221.4	221.2	233.6	249.8	(+ 6.9%)
Japan Key Technology Centre	25.0	26.0	26.0	26.0	(+ 0.0%)
Basic technologies for future industries	6.0	6.4	6.8	7.5	(+ 9.2%)
Large-scale industrial projects	15.1	13.6	13.9	14.1	(+ 13.0%)
Sunshine project	44.1	37.8	27.1	27.5	(+ 1.2%)
Moonlight project	11.4	9.7	10.7	11.6	(+ 7.8%)
Unmanned space platform	0.4	0.5	4.5	5.3	(+ 17.5%)
Superconductivity	NA	2.7	3.8	5.0	(+ 29.3%)
Fifth-generation computer	5.6	5.7	6.5	7.0	(+ 7.6%)
Global environment	—	—	1.0	6.3	(+530.0%)
<i>Ministry of Health and Welfare</i>					
Total R&D budget	39.8	44.1	48.4	51.2	(+ 5.9%)
AIDS	0.1	1.2	2.1	2.2	(+ 1.6%)
Cancer (10-year project)	7.6	8.7	10.7	12.0 *	(+ 11.7%)

NA, equivalent budget not available.

*Includes budgets for Ministry of Education, Science and Culture (¥2.1 thousand million) and Science and Technology (¥7.9 thousand million)

US plan just a memory

San Francisco

THE demise last week of US Memories Inc., a unique and controversial semiconductor manufacturing joint venture, has ended a major attempt by US companies to combat Japanese domination of the critical memory chip arena. The question now is whether it was the last chance, but the venture's chief critic, T.J. Rodgers of Cypress Semiconductor Corporation, still proclaims that the US will "blow away" its Japanese rival early in the next century.

Disagreement centres on whether the combination of improvements in US manufacturing quality — up to now one of the major strengths of the Japanese competition — and the traditional US strength in pioneering new technologies will give the United States a chance to get back on top. Those who helped set up US Memories last June think it will not, and argue that America's only hope is a massive collaborative effort by a consortium of companies. The collaboration was established in the wake of a dramatic shortage of DRAM semiconductors which had caused prices to soar.

DRAMs, which are central to a variety of electronic products, were first developed in the United States by Intel Corporation. But since 1980, the US market share has fallen from 60 per cent to about 15 per cent. Japanese companies now account for nearly 80 per cent of the \$9,400 million market, according to the consulting firm Dataquest.

American chipmakers claim that the Japanese forced them out of the business by dumping DRAMs at artificially low prices. To protect themselves against future shortages and the chance of unfair competition, American buyers wanted another US source. But given the extraordinarily high cost of getting into the business, an estimated \$1,000 million, they wanted to spread the risk. That provided the logic for the establishment of US Memories.

Backers of US Memories included several major semiconductor manufacturers: IBM, Hewlett-Packard, Intel, Digital Equipment Corporation, National Semiconductor, LSI Logic and Advanced Micro Devices. The original plan called for them, and the other supporters that were expected to jump on board, to put up at least half the \$1,000 million needed to establish the new venture as a manufacturer of 4-megabit DRAMs, the next generation of memory chips.

But last week, US Memories president Sanford Kane announced that the effort was dead long before it produced a single product. Several factors marred the effort. The new venture was in some ways similar to SEMATECH, a Texas-based consortium that receives government and

industry funds for research and development into semiconductors, including DRAMS. But unlike SEMATECH, US Memories would have manufactured DRAMS and so may have violated anti-trust law.

From the beginning, critics let it be known that there might be an anti-trust battle. But it was lack of finance that really proved US Memories' undoing. With the DRAM shortage ending, many potential supporters of the venture felt US Memories was no longer vital. At the end of the day, not enough backers were left to make the venture possible.

The demise of US Memories was also a personal victory for Rodgers, a leading critic of the venture and chief executive officer of a highly successful semiconductor company. He argues that the effort smacked of a cartel and that the main problem with US chipmakers in the past was their management of manufacturing.

Rodgers claims that US manufacturers have now so improved yields and quality that they have actually held their own against the Japanese in recent years when the effects of the slumping US dollar are taken into account. That sets the stage for future competition based on innovation and new technologies, areas where, he proclaims, "the Japanese are well aware that we will blow them away."

Not everyone agrees. "Let me tell you, we haven't caught up," says Charles Ferguson, a postdoctoral research associate at the Massachusetts Institute of Technology and author of an August 1989 report on the DRAM industry. Ferguson says that without a US Memories, Americans never will catch up. Start-up costs are so staggering, no company could take on the effort alone. At the same time, prices could tumble, destroying profits. And the government is offering little support in developing a comprehensive national strategy — a shortcoming also noted in a report last November by the National Advisory Committee on Semiconductors.

Even if a company could get established, the odds against success are nearly overwhelming. Ferguson notes that Japan is not only ahead of the United States in facility design, but its workers are much better trained.

US Memories mitigated some of these factors, Ferguson says. He believes the decision by many companies not to back the venture is "extremely shortsighted", and adds, "I think they're going to pay dearly for it".

Andrew Procassini, president of the Semiconductor Industry Association, agrees. He admits the venture might have cost members profits for several years. But he argues that it offered the benefit of supporting the entire semiconductor industry's infrastructure.

Robert Buderl

Cash and teamwork the answer if US to catch up

Tokyo

COMMENTING on the collapse of US Memories, a Japanese expert on the semiconductor industry says that US companies will have to "step on the gas" and "forget about profits" if they want to catch up with Japan.

Hajime Karatsu of Tokai University, a former managing director of Matsushita Communication Industrial Co., is a well-known and outspoken critic of the US semiconductor industry. His opinions appear to agree both with those of Rodgers (see above) concerning the need for improved production practices, and with supporters of US Memories who argue for massive, long-term investment.

Karatsu says he is "disappointed" that consortia such as US Memories are failing but that it is "quite natural" because even in Japan several consortia have shown poor results. He says that one of the problems is that movements in Washington to catch up with Japan are often "biased and political" and guided by people with little practical knowledge of semiconductor production.

Karatsu was invited by the US Department of Defense to point out differences between the US and Japanese industries.

He believes that much of what he said then still applies today. One thing that is needed, he says, is for US industry to learn how to fight against errors, such as the supply of substandard materials and the breakdown of machines. To overcome these errors, all employees from the shop floor to management must learn to work "hand in hand". These conditions prevail in Japan and can be found in the factories of US semiconductor manufacturers in Japan, such as Texas Instruments. But in the United States "individualism" is a handicap in the "war against errors".

Karatsu says he has visited Austin, Texas, many times and has been awarded honorary citizenship for his support of Sematech. But he says US engineers at Sematech want to be independent and tend not to exchange information with each other. If Sematech is to succeed, they will have to learn to share.

The other major problem for the United States lies at the corporate level. US companies will have to invest "much more money" and neglect short-term profits. If not, the distance from Japanese companies will get "further and further", Karatsu says.

David Swinbanks

Fight over Maryland monkeys

Washington

IN a move that suggests the government is preparing to take a more aggressive line in defence of research on animals, government-funded scientists last week experimented on and killed one of the 'Silver Spring monkeys', a group of monkeys that has been the focus of animal activist efforts to improve laboratory conditions.

The monkeys became the *cause célèbre* of the animal rights movement in 1981, when 17 of them were rescued by police from a laboratory in Silver Spring, Maryland, where they had been kept in inhumane conditions. More than 300 congressmen and senators have since called for the monkeys to be retired to a private sanctuary. The controversy over the treatment of the monkeys has seriously damaged relations between the National Institutes of Health (NIH) and Congress, and has been described by the former NIH director James Wyngaarden as the most enervating case of his tenure.

Edward Taub, the researcher who carried out the initial experiments on the animals at the Silver-Spring-based Institutes for Behavioral Research, was arrested in 1981 and convicted on charges of cruelty, although the ruling was later overturned on a technicality. Taub's NIH grant was subsequently withdrawn and the animals were moved to the Delta Regional Primate Research Center at Tulane University.

'Billy', a crab-eating macaque with two amputated forearms, serious spinal deformity and other injuries, was killed on 14 January after four hours of neurological experiments. The operation came only hours after a US district judge lifted a temporary restraining order on the experiment. The restraining order had been obtained two days earlier by animal rights activists determined to stop NIH from killing any of the Silver Spring monkeys. But in a Sunday morning emergency hearing, NIH officials successfully argued that it was imperative that the research be done before Billy died of illness. In part because of the effects of his deformities, Billy had stopped eating and had lost 10 per cent of his body weight.

Animal protection activists lambasted the experiment, accusing NIH of deceit for repeatedly promising that no "invasive procedures" would be performed on the primates. The Physicians Committee for Responsible Medicine (PCRM) filed a complaint with the Department of Health and Human Service's Office of Scientific Integrity Review last week, calling for an investigation into the research and charging William Raub, the acting director of the National Institutes of Health (NIH) and a party to the decision to experiment on the animal, with "gross scientific mis-

conduct." Peter Gerone, principal investigator for the experiment and director of the Delta Center, says that the operation did not violate the "spirit of the agreement" to do no invasive research. "It's not invasive in the sense that we're going to [operate on] the animal and then it's going to regain consciousness", he says.

Although animal activist groups have petitioned NIH to release the monkeys to the care of either of two animal sanctuaries, NIH have so far refused. Agency officials have stated that they do not believe that the proposed sanctuaries can properly care for the monkeys. NIH have meanwhile drafted several proposals for further behavioural and experimental research on the monkeys. Agency officials hope to operate on the rest of the monkeys as their health deteriorates.

Critics say that the monkeys are unsuitable for research and that NIH has concocted the research proposals to avoid turning the animals over to animal protection groups. "It appears that releasing the animals has been viewed by NIH officials as tantamount to surrender", says Neal Barnard, president of PCRM. By experimenting on the animals before they die, he says, NIH hopes to prove that the monkeys did indeed have research value. "NIH needs justification for having kept them so long," Barnard says.

The experiment conducted on Billy sought to determine how the brain reorganizes itself to adapt to the loss of input from nerves elsewhere in the body. Considerable research exists indicating that portions of the brain rendered useless by a nerve injury gradually take over other functions. Further research on the Silver Spring monkeys could help humans with similar injuries, the NIH scientists claim.

But other researchers say the experiment on Billy was unnecessary and poorly planned. "To represent this as science is an affront to science itself," Daniel Robinson, chairman of the Georgetown University psychology department, said at a press conference called by several animal-protection groups. The five-page NIH experimental protocol includes no control group, has no hypothesis or references, and apparently does not take into consideration the varying — and largely undocumented — experience of the animals in the past eight years.

Carrie Walters, a former University of Vermont neurosurgeon now in private practice, says that very little of what can be learned from the Silver Spring monkeys has a human parallel. "You never see a human patient with a totally deafferented limb", Walters says, referring to the condition where all the nerves to a limb have been severed. Because the animals have been isolated for much of the

past eight years and their behaviour was not closely monitored — two important factors in brain adaptation — further research on the animals is likely to be misleading, Walters says. "Bad data is worse than no data", she adds.

The animal-protection advocates said that they would return to the courts to stop further experiments and euthanasia of the seven remaining Silver Spring monkeys.

The latest case also seems likely to renew congressional concern about NIH animal research standards. Roger Galvin, a lawyer for Representative Tom Lantos (Democrat, California) said that the congressman intends to request a congressional oversight hearing on the Delta centre and NIH animal research.

G. Christopher Anderson

AIDS CONFERENCE

European Parliament backs boycott

Paris

THE European Parliament last week adopted a resolution calling for a boycott of the Sixth International Conference on AIDS, due to be held in San Francisco in June. The move is a protest against current US immigration policy which normally denies entry to aliens found positive for human immunodeficiency virus (HIV). For the AIDS conference, a form of *'laissez-passer'* has been proposed by the US authorities, for the duration of the conference, but it requires HIV carriers to register their names.

The resolution, proposed by former French health minister Leon Schwartzberg on behalf of the Socialist Group, was adopted almost unanimously on a show of hands. It calls on the US government to "abolish this discriminatory measure", while also encouraging scientists in the European Communities not to participate in the conference. It proposes that the organizers "transfer this conference to a country where such discrimination does not take place".

At the end of last year, several non-governmental organizations, including the League of Red Cross Societies, decided not to attend the June conference. Jonathan Mann, director of the World Health Organisation's AIDS Committee (WHO), has also put pressure on the US authorities to change their immigration policy. But WHO will soon have to decide whether to withdraw its patronage of the conference, because the current restrictions contravene a declaration it adopted in January last year defending the free movement of HIV carriers.

Following the vote in the European Parliament, the French Health Minister, Claude Evin, sent a personal letter to Mann, calling upon WHO to withdraw from the conference.

Peter Coles

New law needs changes made

Munich

A draft 'framework law' regulating genetic engineering in West Germany was attacked by environmentalists as not being tough enough at a parliamentary hearing in Bonn last week. But members of the conservative majority, while admitting the need for some changes, remain confident that the law will be passed later this year.

The government must act quickly because a November 1989 court decision forbids the use of genetic engineering in industry until there is a legal basis for regulating the technology. The ruling forced Hoechst AG to stop construction of a nearly completed facility.

Adding to the pressure is the knowledge that a majority of the *Länder* (states), which are represented in the upper house of parliament, must approve the law for it to take effect. The conservative coalition of Helmut Kohl has a majority in the upper house at present, but that could change in an election in Lower Saxony on 13 May. If the law cannot be passed by then, the opposition parties may attempt to make it more restrictive.

The acrimonious three-day hearing was dominated by activists who claim that genetic engineering is an inherently hazardous technology. The text of the law seems to concur — the first paragraph states that a law is needed in order to protect "life, property and the environment . . . from the dangers of genetic engineering . . .". But the biotechnology industry supported the law at the hearing, and basic research organizations said they could live with it, if there must be one, as long as a few changes are made.

Not all the changes are likely to please industry. The biggest change expected will give the *Länder*, not Bonn, the authority to grant licences to basic research laboratories as well as to factories that use recombinant DNA technology.

Ernst-Ludwig Winnacker, vice-president of the Deutsche Forschungsgemeinschaft, said that if this happens, he fears uneven treatment of licence applications in the different *Länder*. He questions whether there are enough qualified experts in West Germany to fill all the review committees. But Winnacker hopes that the *Länder* will accept decisions made by the Central Commission for Biological Safety (ZKBS in German), which currently issues opinions on licensing matters.

Almost certain to change is the composition of ZKBS. ZKBS is composed of eight biologists and four representatives of unions, industry, environmental and research organizations. But parliamentarians of all parties seem to feel that the committee contains too many 'experts' who will inevitably favour genetic

engineering. Winnacker criticized this thinking, saying that the government will retain the power to overrule ZKBS so that there is no need to pack it with people without a scientific background.

The West German law must fit with an European Communities (EC) directive on genetic engineering that will take effect next year. The EC directive foresees just two safety classifications for genetically manipulated microorganisms used in confinement, whereas the West German law foresees four levels.

Winnacker would like to see the West German law adapted to the EC directive in other ways too, for example by removing from regulation certain nonrecombinant methods that lead to gene transfer, such as mutagenesis, hybridoma formation and protoplast fusion.

Although he was glad that the release of genetically engineered organisms into the environment will be allowed under the new law, Winnacker warned that the law does not distinguish between different types of release. Certain harmless forms of release, such as the sale of transgenic

UK UNIVERSITIES

Equipment lack documented

London

AN injection of £259 million is needed to meet the demand for research equipment in UK universities, according to a report to be published shortly by the Advisory Board for the Research Councils (ABRC). The report, compiled by Luke Georgiou, Peter Halfpenny and Susan Hinder of the University of Manchester, also shows that researchers in more than 80 per cent of university departments feel that critical experiments are either prevented, or delayed, because of lack of equipment.

The survey took the form of detailed questionnaires, issued to university departments in 1988, and looked at equipment in the £10,000 to £1 million price range. Researchers had to explain how desired items of equipment were needed.

Many of the required items of equipment were more expensive than the average cost of equipment surveyed, suggesting that there is a particular problem in funding for more expensive equipment, such as spectrometers and nuclear magnetic resonance (NMR) equipment.

British academics also believe their laboratories fall below top international standards. Eighty-one per cent of departments felt they were well-equipped by UK standards, but only 22 per cent thought this was the case in international terms. British equipment was markedly older than that in US universities surveyed by the National Institutes of Health and the

animals and plants, must be specifically allowed by the law, he said.

One improvement, from the point of view of industry, will come with the introduction of a one-step licensing procedure for factories that use genetic-engineering technology. All the issues raised by the use of genetic engineering will be addressed in a single public hearing. Once this is over, the company can use the factory in question for a variety of techniques without requiring a new hearing for each microbial strain.

Coalition member Heinrich Seesing (Christian Democrat), who is a member of the subcommittee responsible for the law, was confident that it could be rewritten and passed in time. "There is a majority in parliament in favour of allowing both genetic engineering and environmental release", he said.

But opposition member Edelgard Bulmahn (Social Democrat), who is also on the subcommittee, warned that trying to push the law through parliament against the will of the opposition would be very unwise. She said that the *Länder* governed by Social Democrats would balk at administering a law they found unsatisfactory.

Steven Dickman

National Science Foundation in the mid-1980s. These surveys judged US facilities inadequate and led to calls for upgrading.

Both Georgiou and Halfpenny believe that more rigorous international comparisons are an important next step in their work. Georgiou says that data from new US surveys, France and West Germany will be re-analysed along the same lines as their own survey.

Apart from a massive injection of cash, the ABRC report identifies greater sharing of equipment between institutions as a potential solution to the problem. The report shows that, on average, equipment lies unused for 26 per cent of the time. Some schemes to promote equipment sharing do already exist. The Science and Engineering Research Council (SERC) runs loan pools for some equipment, and demands wider access as a condition when some expensive items, such as NMR equipment, are awarded to particular institutions. Geoff Richards, of SERC's science division, believes that equipment sharing has an important role to play, but this role is "limited by the demands of the science involved".

Scientists surveyed in the ABRC study were not happy with current arrangements for access to nominally shared equipment. Hinder, one of the authors, believes that a more formalized system for sharing is needed.

Peter Aldhous

RADIOISOTOPE HANDLING

Scandal at Tokyo hospital

Tokyo

FOLLOWING a tip-off from a newspaper, officials of Japan's Science and Technology Agency (STA) this week began digging up the grounds of the nation's leading university hospital looking for radioisotopic waste that may have been dumped illegally. The search is just one incident in a scandal at the Tokyo University hospital that has drawn attention to problems associated with the explosive growth in use of radioisotopes for research and medical purposes in Japan.

The scandal erupted two weeks ago when newspapers reported that an STA inspection of Tokyo University Hospital in November had revealed 30 serious defects in the handling of radioisotopes by the hospital, including failure to keep complete records of the use and disposal of radioisotopes, using radioisotopes in a research laboratory undergoing construction, using one-year's authorized supply of radioisotopes in six months and failure to supervise and educate researchers in the use of isotopes. The director of the hospital has publicly apologized and STA has banned the use of isotopes at the hospital for research purposes until the situation is rectified.

In addition to the mishandling pointed out by the agency, the *Mainichi* newspaper revealed that in dismantling a 20-year old betatron at the hospital last October, researchers were surprised to detect high levels of the dangerous isotope strontium-90. The strontium-90 dates from long-forgotten experiments carried out when the betatron was first built, according to Tetsuhiko Yoshida, director of STA's Radiation Protection Division.

The *Mainichi* also disclosed with photographs that on 7 November, just before the STA inspection, the hospital used part-time labourers to remove drums of radioisotope waste from the basement of the hospital without instructing them about the materials they were handling and without requiring them to wear film badges to determine their exposure to radiation.

This constitutes a violation of Japanese law, according to Yoshida. But the hospital has denied legal responsibility.

The decision to dig up the hospital grounds stems from another claim by the *Mainichi* that one of the hospital's own researchers told the newspaper that radioactive waste, including beakers and flasks, is buried there. In a preliminary survey last week, STA officials detected four 'hot spots' in the courtyard of the hospital with levels of radioactivity 10 to 100 times that of background.

There has been explosive growth in the number of organizations using radioisotopes for research in Japan over the past

30 years, and the amount of untreated (and partially treated) radioisotope waste increased nearly sixfold between 1980 and 1988 to more than 50,000 200-litre drums. But the Atomic Energy Commission is only just beginning to look into possible sites for waste disposal.

Yoshida thinks that the mishandling of radioisotopes at Tokyo University Hospital is a special case. Each year, the agency

ALCOHOL FUEL

Brazil's pumps run dry

São Paulo

BRAZIL's alcohol fuel programme, the largest alternative fuel programme in the world, is in dire straits. The supply of sugar-cane ethanol has failed to keep up with demand and, over the objection of environmentalists, the government is now coming up with a new 'cocktail' fuel that is part methanol, part ethanol and part gasoline. The alternative would be rationing of fuel for the four million Brazilians who own ethanol-driven cars.

Last week, environmentalists tried legal action to block emergency measures to add methanol to the sugar-cane alcohol 'gasohol'. They argue that methanol is highly toxic and that little is known of the emission from methanol-powered cars.

Legal action was not a success and the federal court reversed an earlier decision favouring prohibition of methanol. But the decision holds for only six months and cannot affect states and cities where the local executive or legislative have forbidden ethanol use. The decision calls for protection standards to be established by the Health Ministry and for a media campaign on the risks of the new 'cocktail' fuel. Besides being easier to ignite, the new fuel can pose an additional hazard to a well-established Brazilian habit — siphoning fuel from tanks. The siphoners are both careless people who forget to fill their tanks and have to borrow from others, and less well-intentioned fuel thieves. Methanol ingestion can lead to blindness and death.

Researchers from the University of São Paulo have presented a report on the possible environmental effects of methanol use as has CETESP, the São Paulo state pollution control agency. Both groups advise that methanol use be coupled with monitoring of environmental effects and a campaign to inform the public of its risks.

CETESP officials are particularly angry with Brazilian environmentalists. Their opposition to the use of alcohol would mean more petrol use, and that would mean more carbon monoxide in the stifling air of major cities such as São Paulo and Rio de Janeiro. Ironically, these were

carries out safety checks at about 400 organizations that use radioisotopes for research purposes but none has shown such a large number of safety violations.

Yoshida sees three major causes for the problems at Tokyo University: the hospital building is old and poorly maintained, the number of technical staff trained in the safe use of radioisotopes is very limited and Tokyo University doctors have a high status in Japanese society and tend to ignore the advice of trained technicians.

David Swinbanks

among the first cities where local authorities banned methanol. But São Paulo's mayor, Luiza Erundinia, changed her mind last week and said she will permit methanol use, provided protection measures are taken.

Brazil's alcohol programme has been suffering from mismanagement and sheer bad luck. The country is the biggest sugar-cane grower and the biggest producer of sugar and ethanol. And yet there is a serious shortage of alcohol.

Several factors are to blame. Instead of meeting government alcohol quotas, sugar-cane planters and distillery owners preferred to produce sugar and export it while prices were good.

Crops were not as large as expected. Government alcohol stocks were allowed to run down because of financial problems at Petrobras, the state oil company which buys alcohol from producers.

Consumption of alcohol has recently risen thanks to the short-lived success of the summer economic plan of 1989, and industry is still geared up to produce more ethanol-powered cars than gasoline-powered ones, although that is changing. In 1985, sales of alcohol cars reached an all-time high of 96 per cent of the total; today it is down to 61 per cent. Brazil now has roughly a third of its 13 million vehicles running on alcohol.

Subsidies to sugar growers and distillers — a powerful lobby ever since colonial times — also distort fuel economics. Alcohol costs less to consumers than its real price, and gasoline costs more. Thanks to this policy, Petrobras has run up a huge deficit since the Proalcool programme began in earnest in the 1980s.

Researchers have also come up with unexpected news about alcohol pollutant emissions. It is true that carbon monoxide and particulate levels are down. But alcohol fuels produce more aldehydes, which react in the atmosphere and create peroxyacetyl nitrate (PAN), a compound that stunts plant growth and causes eye irritation. Aldehydes also help increase ozone levels, one ingredient of smog.

Ricardo Bonalume Neto

UK welcomes help from Japan

Tokyo & London

THE Japanese government is on the verge of investing several million pounds in the United Kingdom for two joint research projects. One was discussed by Prime Minister Toshiki Kaifu and Prime Minister Margaret Thatcher on 12 January during Kaifu's visit to Europe and an agreement is expected to be signed next month. The other is still under negotiation.

The project discussed by Kaifu and Thatcher is the first of a new international research programme funded by the Research and Development Corporation of Japan (JRDC), an affiliate of the Science and Technology Agency (STA). The new programme, which began this fiscal year, is modelled on JRDC's small but successful Exploratory Research for Advanced Technology (ERATO) programme (*Nature* 338, 532; 1989).

The project, to be called the 'Atom Arrangement Design and Control Project', will be carried out at the University of Cambridge and at the Interdisciplinary Research Centre on semiconductor materials at Imperial College, London, in collaboration with the National Research Institute for Metals (NRIM) in Japan. A joint team of UK and Japanese researchers at Imperial College will investigate layer-by-layer growth of semiconductor films by molecular beam epitaxy. Another team at Cambridge will study the microstructure of high performance alloys, for example aluminium alloys used in aircraft turbine

blades.

Final details of the project, which is expected to last five years with a budget of about £10 million, have yet to be settled. But it appears that the United Kingdom will have to produce little or no new funding for the project. Rather, JRDC will accept 'funding in kind' in the form of research facilities and resources as the United Kingdom's 50 per cent contribution. Japan will provide postdoctoral researchers and cash for new equipment.

JRDC initially approached the US National Science Foundation (NSF) last year in the hope of arranging the first joint project. But NSF officials found descriptions of the new programme "vague" and raised a lot of questions. However, a JRDC official in charge of the programme says he is confident a joint project will

soon be started.

JRDC has won a budget of ¥364 million (\$2.6 million) for the programme in fiscal year 1990. ¥280 million will go to the joint project with the United Kingdom and the rest to a new project. Eventually JRDC officials hope to have about four new projects each year, as in the case of ERATO.

The other joint UK-Japanese project, still under negotiation, is on muon-catalysed fusion and is expected to be carried out at the Rutherford Appleton Laboratory in the United Kingdom. Japan will provide a negative muon channel with superconducting magnets and other peripherals for the high-intensity pulsed-muon source at Rutherford. STA has received preliminary funding for the superconducting channel. The total cost of the project over the next seven years is expected to be about £10 million.

David Swinbanks & Peter Aldhous

Still a step from 'Eureka'

Paris

FRENCH companies may be shying away from the Eureka programme for marketable European high-technology innovation because of inefficiency in the way it is run. This is one of the conclusions revealed in an independent 'operational audit', commissioned by the government, of about half of the 127 projects in which French industry has a share.

Nevertheless, the survey, carried out by IDS Consultants, a company based in the Paris suburbs, says that Eureka is "an undeniable success", with 297 active projects. But to make sure that French industry continues to be involved in new proposals, Prime Minister Michel Rocard last week unveiled a series of measures designed to improve efficiency in the selection, financing and follow-up of Eureka projects — the main stumbling-blocks identified by IDS.

According to the survey, projects may still be "very vague" when they are accorded Eureka status; sometimes evaluations are inadequate and may be contested; some of the projects do not have obvious industrial applications (contrary to Eureka's 1985 Hanover Charter) and, sometimes, 'alibi' partners are brought in to conform to the rule of transfrontier collaboration, although their involvement is not essential.

Companies also criticized the way government grants are managed. More than 89 per cent of respondents complained that backing was not guaranteed for the duration of the project. Sixty per cent felt it took too long for the authorities to agree projects and to sign a contract. About half said the sizes of grants promised by the government were vague and noncommittal,

while differences in the ways partner countries finance projects can cause stops and starts. Finally, when there are several sponsors, financial management becomes overcomplicated.

To counter these criticisms, the inter-ministerial Eureka committee, presided over by research minister Hubert Curien, has taken steps to tighten up the administration. More emphasis will be placed on the initial evaluation of proposals, helping companies to define their projects. And once a project is accepted, its evaluation committee will continue to steer developments during its passage to production.

A greater responsibility has also been given to ANVAR (Agence Nationale pour la Valorisation de la Recherche), the governmental agency set up to promote technology innovation. ANVAR will help companies to find foreign partners — one of the main reasons small companies seek Eureka support — and will serve as cashier when there are several sponsors.

The Eureka committee also wants to see greater diversity in the fields of technology covered. So far the most successful projects have been in the electronics sector, such as microchips (JESSI) and high-definition television. In 1990, a push will be made to find new projects in seven target areas: the construction, automobile and railway industries, pharmaceutical and biotechnologies, telecommunications, agriculture and the environment.

Last week, the French Eureka programme also moved out of its probationary period when a new statute established a permanent organization committee, headed by secretary general Henri Guillaume.

Peter Coles

DARPA

Neural network funding in doubt

Washington

A 'miscommunication' between Congress and the Defense Advanced Research Projects Agency (DARPA) could result in a major cutback for neural network research in 1990. Congress last year stipulated that the Defense Department programme was to get \$20 million this year from 'reallocated' 1989 funds. But when DARPA officials prepared to distribute the money earlier this month, it became quickly apparent that there were no funds to be had, says programme director Barbara Yoon. The agency is looking for alternative funding from elsewhere in the Defense Department, but Yoon says that "it's safe to say that the programme will have to be scaled back" for the year. She emphasizes, however, that the cutback should not be interpreted as a lack of enthusiasm within DARPA for neural network research.

Yoon could not say how many grants have already been approved, or how many will be cancelled. G. Christopher Anderson

Shift away from NASA

Washington

THE White House has decided to look outside NASA (the National Aeronautics and Space Administration) for advice on the proposed manned mission to Mars, a shift that may be the first sign of a growing power struggle between the space agency, the vice-president-led National Space Council, the aerospace industry and government research laboratories over who shall take part in and plan the massive two-decade project.

In a speech to the American Astronomical Society in Washington earlier this month, Vice President Dan Quayle announced that he had asked the Aerospace Industries Association and the National Research Council independently to review proposals for going to the Moon and Mars in the next century.

"The hypothesis, which remains to be demonstrated, is that there are a lot of ideas out there that NASA has rejected because of institutional bias", says George Washington University space policy analyst John Logsdon, a Space Council adviser.

The decision indicates a diminishing White House confidence in NASA. Late last year, the agency submitted to the Space Council a preliminary plan outlining five similar scenarios for the mission, ranging from 18 to 25 years in duration. But critics of the plan say that it reflects NASA's commitments and allegiances to existing projects and relationships more than innovative thinking about long-distance space travel. All five NASA options, for example, depend heavily on the Space Station, a troubled project that

is in dire need of a compelling mission such as the Mars project.

"To no one's surprise, NASA came up with a plan that was essentially aimed at keeping the contractors employed", says John Pike, space policy analyst for the Federation of American Scientists. The Space Council's decision to widen the planning process is a sign that the NASA plan is "dead in the water", he says.

Some members of the aerospace industry are privately hopeful that Pike's interpretation is correct. Although many aerospace companies have links with NASA, the Defense Department has historically been a far richer source of funding. With major budget cuts planned in defence programmes, however, industry executives are eager to find a non-defence market for their high technology. One solution would be to strengthen their ties with NASA. But NASA loyalties are well established.

The Space Council's decision may offer the aerospace industry a good chance to get in on the ground floor. The Aerospace Industries Association says it will look primarily among its members for new ideas to bring to the Space Council. A report is due by the end of March.

For the defence-orientated national laboratories such as the Department of Energy's Lawrence Livermore and Los Alamos facilities, the thawing of the cold war has likewise left many laboratory scientists enviously eyeing the Mars mission. Both laboratories are intimately involved in the Strategic Defense Initiative, for which funding seems set to plummet.

Livermore founder Edward Teller and his protégé Lowell Wood have both spent time talking to the Space Council about proposals for Mars. Quayle is said to have been impressed by one idea of Wood's for using inflatable Kevlar-covered balloons for space travel and both Mars and Moon bases. The double-walled 5 metre x 15 metre structures could be stowed compactly until needed, Wood says, greatly simplifying the process of establishing living quarters or storage space. Although Wood acknowledges that the technique carries some risk — the balloons might pop — he claims the technique could cut the project's cost by an order of magnitude.

That idea and others gathered by NASA will be the focus of the NRC study. A panel headed by former White House science adviser Guyford Stever will prepare a report by the end of February. Some speculate that the hurried timetable is to put together a viable plan in time for the planned US-Soviet summit this spring.

Gorbachev is expected to suggest going to Mars together. With a plan that has come from the National Research Council, the Aerospace Industries Association and NASA, Bush may finally be able to agree.

G. Christopher Anderson

PATENTS

Cetus battles with Du Pont

Washington

WITH the recent issue of a broad patent covering *Taq* DNA polymerase to Cetus, the California-based biotechnology corporation looks set to dominate the lucrative US market for PCR (polymerase chain reaction) based products. Cetus has been quick to protect its patent rights to the PCR technique, which was first fully described only five years ago. This is the fourth US patent awarded to Cetus for PCR and related enzyme technology — another 40 are pending.

The US patent for *Taq* DNA polymerase — the thermostable enzyme that catalyses the DNA amplification technique PCR — was issued on 26 December. It covers the full-length enzyme, as purified from the bacterium *Thermus aquaticus* or that produced by genetic engineering. Peter Staple, associate general counsel for Cetus, says the *Taq* DNA polymerase patent effectively bars anyone from "making, using or selling" the enzyme in the United States.

News of the patent will alarm the many US biotechnology companies that manufacture *Taq* DNA polymerase and *Taq*-based products. Staple says that Cetus has already been approached by several companies keen to produce under licence and, although this may be permitted for "certain" applications, requests would be reviewed case by case.

Cetus will not disclose its revenue from *Taq* DNA polymerase, which is marketed through the Perkin-Elmer joint venture, but says it is one of its largest selling molecular-biology reagents. The joint venture with Perkin-Elmer, which includes a thermal cycler used to automate PCR, should earn Cetus \$60 million a year by

the mid-1990s.

While the latest patent will strengthen Cetus's overall position regarding PCR, the company is still engaged in a bitter dispute with the biotechnology giant, Du Pont. Last August, Du Pont filed a law suit against Cetus challenging the validity of its two key PCR patents which cover the PCR process and the detection of PCR products. Du Pont cites two papers suggesting a possible method for DNA amplification published more than ten years before Cetus first described PCR (see *Nature* 342, 9; 1989).

Du Pont has made a "substantial investment" in the development of PCR-based kits for the diagnosis of human diseases. It has been unable to obtain a licence from Cetus, which has licensed all diagnostic uses of PCR to Hoffmann-La Roche. The suit is seen as a pre-emptive strike by Du Pont which is under "reasonable apprehension" of being sued by Cetus for infringement of the two PCR patents.

Hoffmann-La Roche, which stands to gain an estimated \$600 million a year by the mid-1990s from the sale of PCR-based kits and assays, has petitioned the patent office for a "re-examination" of the two patents. Staple says this legal provision determines whether there is a question of patentability in claims of prior art, and usually results in an upholding of the original patent.

The court case between Du Pont and Cetus is set for 7 August. But Du Pont's plans to go ahead with the launch of its first PCR-based product, an HIV-1 test kit, at the end of this month is likely to spark a new wave of controversy and yet more litigation.

Diane Gershon

Antecedents of a Nobel prize

SIR—The award of the 1989 Nobel Prize in Medicine and Physiology to Drs J. Michael Bishop and Harold E. Varmus has ended up in a controversy raised by Dr Dominique Stéhelin. As an author of one of the two papers and one who worked for the first six months on this project, I am well placed to describe the events leading to that masterpiece work. I record here the relevant chronology as I know it so that the scientific community can judge that the prize was properly awarded.

I joined Bishop's laboratory on 4 January 1973 as an assistant research microbiologist. Stéhelin had preceded me by a few months and was working on characterizing polysome-associated virus-specific mRNA. Very soon, Stéhelin and I became close friends and began to exchange views on a variety of issues, ranging from science to politics. He was somewhat frustrated because his project was not going well and because of personal problems.

Before joining the laboratory, I had set my mind on working on the various DNA intermediates in retrovirus-infected virus cells because of the spectacular discovery of reverse transcriptase by Mizutani and Temin, and by Baltimore, in 1970 (Mizutani did not get the prize, Temin did). When I arrived in the laboratory, Bishop suggested that I should talk to all the postdoctoral fellows and technicians to familiarize myself with various projects. After several discussions, I made up my mind to work on DNA intermediates, as nobody else was actively pursuing that topic. Varmus himself was working on integration by network formation followed by reassociation kinetics. He advised me on various tissue culture and molecular biology techniques, and Nancy Quintrell, a research assistant, provided me with valuable information on preparing probes and so on.

After a week, I had a meeting with Bishop and Varmus, at which Varmus suggested that I should first start by preparing the *src* probe. He outlined the rationale, which was based on the original observation by Peter Duesberg and Peter Vogt that transformation-defective variants lack a sequence of about 1,000 to 1,500 nucleotides. It sounded to me very interesting and I therefore set out to do the experiments.

I prepared radiolabelled cDNA probes from wild-type virus and hybridized them to the 70S RNA from transformation-defective (td) virus. I began to notice differences in the extent of hybridization between them and therefore pursued this observation vigorously for three or four months. As my heart was set on doing experiments with viral DNA intermediates, I began those experiments simultaneously. In about two or three months, I

obtained evidence that the first steps of reverse transcription occur in the cytoplasm. We were all excited by what seemed to be a seminal observation, but I had to prove that the DNA we detected in the cytoplasm was not due to leakage from nuclei. After these initial observations, Varmus started enucleating the cells to show conclusively that DNA synthesis occurs in the cytoplasm.

In the next few months, we concentrated on this project, which resulted by the end of 1973 in the discovery of supercoiled DNA. As a result of this diversion, I had to slow down the work on *src*, and was only occasionally able to work on that project. One day in July 1973, Bishop came to me and said, "Ram, since your project on DNA synthesis is going so well, do you mind if Dominique continues this project; but I want to assure you that you are part of this project as well".

I readily accepted the proposal, since Stéhelin was (and is) a good friend, and his original project was not going well. I handed over my notebook to Stéhelin. He started working on this project at the end of July or August 1973, while I pursued the studies on DNA intermediates. By the end of 1973, we found supercoiled DNA, and the work on this was presented in 1974 at a symposium held at Rutgers University and at the Cold Spring Harbor meeting and later published in *Nature* in 1975.

I continued to attend the weekly meetings with Bishop, Varmus and Stéhelin until the middle of 1974, but discontinued later because of my own projects and also because of my intensive effort at job hunting. Stéhelin worked hard, and successfully prepared the *src* probe. Once the probe was defined, it was logical to evaluate it on all the DNAs available.

In my view, Bishop and Varmus deserve the Nobel prize. Several others such as Peter Duesberg, Peter Vogt, H. Hanafusa, Jeff Cooper and R. A. Weinberg could have been included, because the first three laid down fundamental work which made possible the oncogene work on retroviruses and the last two contributed to identifying cellular oncogenes not present in retroviruses.

In retrospect, it seems to me that if I had not obtained positive results, there would have been no oncogene project. In many laboratories, many students and research fellows begin work on new projects, but not many succeed. Initial failure of any project results, in general, in its abandonment. Very few continue on a project when several experiments fail. That was what happened to Stéhelin's first project and therefore, to my mind, fortuitous circumstances paved the way for his involvement in the oncogene project.

Therefore, as the first participant in the early work, in the 1970s, which led to the

award of the 1989 prize, I believe that the Nobel committee acted prudently in the selection of Bishop and Varmus. I hope that the Nobel committees will continue to base their decisions on scientific merit, ingenuity and the impact of discoveries and ignore the publicity and politics which develop following some discoveries, such as that of the AIDS virus, which is but one among many such discoveries and controversies of the past decade.

RAMAREDDY V. GUNTAKA

Department of Molecular Microbiology
and Immunology,
University of Missouri-Columbia,
Columbia, Missouri 65212, USA

SIR—Under the heading "Conduct unbecoming", *Nature*¹ says that "As all agree, Stéhelin was the main pair of hands behind the experiments that first showed that the oncogenes of tumour viruses are stolen and corrupted versions of genes from the cells. . .". The latter half of this statement is incorrect for the following reasons.

The first demonstration of the cellular origin of viral oncogenes was published from entirely different laboratories three years before the paper by Stéhelin *et al.*² appeared in 1976. Scolnick *et al.*¹⁰, cited in Stéhelin *et al.*'s 1976 paper, and Tsuchida, Gilden and Hatanaka⁴ observed in 1973 and 1974 respectively that the *ras* sequences, then called *src*, of Harvey and Kirsten sarcoma viruses were from cellular genes. The title of a paper by Scolnick and Parks in 1974 is "A second murine Type C virus with rat genetic information"³ and that of a paper by Tsuchida, Gilden and Hatanaka is "Sarcoma virus related RNA sequences in normal rat cells"⁴. The protocol of these experiments was exactly the same as that used later by Stéhelin *et al.* and so was the conceptual basis, namely to test Huebner's and Todaro's oncogene hypothesis of 1969⁵ with the then newly discovered viral oncogenes. The oncogene hypothesis postulated the existence of switched-off cancer genes in normal cells. The only technical distinction between the original experiments with the Harvey and Kirsten viruses and the later ones by Stéhelin *et al.* with the Rous sarcoma virus was that, in addition to sequences from the cell, the Harvey and Kirsten viruses had also 'stolen' sequences from an endogenous rat retrovirus. This was painstakingly and convincingly sorted out by the Scolnick group in subsequent years.

Further, there is no mention in Stéhelin *et al.*'s paper² that the sequences transduced had been 'corrupted' cellular sequences. Instead, these sequences were called viral homologs of cellular genes in this² and subsequent studies⁶. Indeed, on the basis of this homology, the cellular progenitors of viral *onc* genes were proposed to be switched-off cellular

oncogenes, abbreviated to this day as *c-oncs*, that would function as cancer genes by increased dosage⁶. Hence the term oncogene activation⁶. But subsequent work, mostly by others, has shown that viral *onc* genes and cellular progenitors have different functions because they have different structures⁷⁻⁹. So far, not a single cellular progenitor of a viral *onc* gene has been proved to be a cancer gene, except if it had been altered ('corrupted') by recombination with a retrovirus⁷⁻⁹. Yet it seems that the appeal of this hypothesis is the probable basis for the current Nobel prize.

All the points I make here had been made before the current controversy, first in *Nature* and later in other journals⁷⁻⁹ and have not so far been challenged.

PETER DUESBERG

University of California, Berkeley,
Department of Molecular and
Cell Biology,
Stanley Hall,
Berkeley, California 94720, USA

1. *Nature* **342**, 328 (1989).
2. Stéhelin, D. et al. *Nature* **260**, 170-173 (1976).
3. Scolnick, E.M. & Parks, W.P. *J. Virol.* **13**, 1211-1219 (1974).
4. Tsuchida, N., Gilden, R.V. & Hatanaka, M. *Proc. natn. Acad. Sci. U.S.A.* **71**, 4503-4507 (1974).
5. Huebner, R.J., & Todaro, G. *Proc. natn. Acad. Sci. U.S.A.* **64**, 1087-1094 (1969).
6. Bishop, M. & Varmus, H.E. in *RNA Tumor Viruses*, Weiss, R., Teich, N., Varmus, H. & Coffin, J. (eds) 999-1108 (Cold Spring Harbor Laboratory, New York, 1982).
7. Duesberg, P.H. *Nature* **304**, 219-226 (1983).
8. Duesberg, P.H. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2117 (1987).
9. Duesberg, P.H., Goodrich, D. & Zhou, R.-P. *Ann. N.Y. Acad. Sci.* **567**, 259-273 (1989).
10. Scolnick, E.M. et al. *J. Virol.* **12**, 458-463 (1973).

Motive power

SIR—Virginia Trimble clearly establishes (*Nature* **342**, 11; 1989) that graduates of more prestigious universities publish for a longer professional period than those of other institutions. No one will want to squabble with this finding. But the conclusion that she draws, that "it still pays to go to the most prestigious graduate institution", is based on an assumption that is congenial to a physicist used to classes of identical objects. It is the assumption that the samples of students in the two kinds of institutions are equivalent and that their lifelong behaviour is the result of their institutional association. This ignores, however, the fact that only one in seven or even one in 10 of the applicants to a prestigious institution is accepted, while in less prestigious institutions the selection is far less severe.

One of the criteria on which this drastic selection is based is an indication of high motivation of the applicant. This criterion is of great importance because many studies have shown that high motivation, all else being equal, is usually more important for high achievement than mere intelligence. In other words, there is much to indicate that the continuing high moti-

vation of graduates from prestigious institutions is due to the fact that they represent a sample selected for high motivation.

This conclusion does not deny the additional factor of an intellectual climate in the prestigious institution that favours continuing productivity nor other aspects of the environment (such as lower teaching load) that also favour high productivity. Yet all the evidence indicates that continuing high motivation is not primarily the effect of institutional association.

ERNST MAYR

Museum of Comparative Zoology,
Harvard University,
Cambridge, Massachusetts 01138, USA

Nonrandom uses

SIR—The word "random" is sometimes used in *Nature* and other publications where another term such as "chaotic", "unpredictable", "uncertain", "arbitrary" or "undetermined" should be used. The use of "random" to describe a process, behaviour or physical system should be reserved for cases where an author can prove the system is random.

The primary definition of "random" is "proceeding, made, or occurring without definite aim, reason, or pattern"¹. "Apparently random fluctuations"² would be better expressed as "unpredictable fluctuations" as there are "structures" and "regularity".

When "random" is used to denote a selection process³ or simulation method, close attention must be paid to the implementation of the randomizing process. That the chaotic orbits of a three-body system would upon ejection of a member "randomize the spin orientation"⁴ is neither demonstrated in the paper nor an expected result of the ejection. Instead, the spin orientation is deterministic and yet unpredictable. The reason and pattern for the spin orientation are known; it is the limits of observation and the limits of computability that prohibit prediction.

In a statistical sense, "random" is "of or characterizing a process of selection in which each item of a set has an equal probability of being chosen¹ and is properly used in *Nature* **342**, 128 (1989) when speaking of a "random distribution of orientations" as a metric.

Even the use of "random" in statistical analysis and computer science must be viewed critically⁵. Computers are used to generate sequences representing random selections using deterministic algorithms. Some common computer randomization methods produce less than exemplary results. These algorithms produce reproducible sequences that have qualities that emulate "random" behaviour (if such a thing as random behaviour exists). These computer-generated sequences are better termed "pseudo-random".

If system history influences succeeding

states, then we cannot consider it random merely because we do not know the aim, reason or pattern. In dynamics we must carefully consider evidence such as power spectral distribution, trajectory, degrees of freedom, rate of divergence and dimensionality when branding a system "chaotic". The same burden of proof should apply to those who report a system to be "random".

THOMAS L. OCHS

US Bureau of Mines,
1450 Queen Avenue, SW,
Albany, Oregon 97321, USA

1. *Random House Directory of the English Language* 2nd Edn (1987).
2. James, I.N., & James, P.M. *Nature* **342**, 53 (1989).
3. Shulman, S. *Nature* **342**, 4 (1989).
4. Ables, J.G. et al. *Nature* **342**, 161 (1989).
5. Park, S. K. & Miller, Keith W. *Random Number Generators, Good Ones Are Hard To Find, Communications of the ACM* **31**, 1192 - 1201 (1986).

Embryo research

SIR—Public attitudes to fertilization and embryology have been shown to be considerably pliable. This is well illustrated in the Warnock Report by the issue of artificial insemination by husband (AIH) and by donor (AID). The Archbishop of Canterbury in 1948 was highly critical of the practice of AID, though not of AIH, recommending that it should become a criminal offence. But no action was taken.

In 1960, the Feversham Committee, set up by the government to consider AI, reported; it considered that AIH was an acceptable form of treatment for some couples, but believed that most people within both society and the medical profession were opposed to the practice of AID. It concluded that AID was an undesirable practice, strongly to be discouraged. Since 1960, the practice of AID has continued to grow. In 1968, the then Minister for Health decided that AIH and AID should be available within the NHS if recommended on medical grounds.

I do not wish to discuss whether such a change has been for the better, but the history of attitudes towards AID illustrates how, in the absence of parliamentary restrictions, the views of a progressive minority may become widely adopted. Although the issues involved in *in vitro* fertilization are not the same, it might be argued that they are of sufficient importance that future changes on such matters as a time limit for experimentation should be subject to parliamentary review. Opinions such as "Why not start with a 14-day limit [on embryo research] imposed by regulation with the intention of relaxing that if the new authority can make a convincing case for doing so, and win public acceptance for its plans?" (*Nature* **342**, 461; 1989) fuel the fears of those opposed to any advance.

JONATHAN J. EWBANK

MRC Laboratory of Molecular Biology,
Hills Road, Cambridge CB2 2QH, UK

Sunspots and influenza

SIR—Some years ago, Hope-Simpson¹ pointed out a remarkable coincidence between peaks in the sunspot curve, when solar activity is at a maximum, and the occurrence of influenza pandemics associated with antigenic shifts of the virus. He pointed to five coincidences over the period 1919–68 as shown in the figure below. Using historical evidence on flu-like epidemics as summarized by Beveridge², we extended the comparison before 1919 as shown³:

Sunspot maximum	Date of pandemic
1761	1761–62
1767	1767
1778	1775–76
	1781–82
1787	1788–89
1804	1800–02
1830	1830–33
1837	1836–37
1848	1847–48
	1850–51
1860	1857–58
1870	1873–75
1893	1889–90

The first column ignores the abnormally low sunspot maxima in the years 1816, 1883 and 1905, and the second column includes both “certain” and “possible” pandemics in the judgement of Beveridge². Although the correspondences between the figure and our table are clearly not precise, they do add credence to the speculation that solar activity and influenza activity may have a causal link. The two phenomena, which have irregular periods (with an average period of 11 years), appear to have kept in step over some 17 cycles.

Past experience has shown that false correlations of phenomena with the sunspot cycle may look good over a few cycles but go seriously adrift after an appreciable number of cycles. This does not happen for the postulated sunspot–flu connection.

Following the broad sunspot peak in 1968–70, the next peak occurred in 1979–80 coinciding closely with the ‘Red Flu’ pandemic of 1978–79 (see figure). The sunspot relative numbers from May to October 1989 are given as follows: 159, 166, 172, 176, 176, 175. The indications are that a peak has been reached in August/September 1989, and that this peak is the second highest on record. The highest recorded sunspot peak which occurred in 1957 coincided with the Asian flu pandemic of 1957–58. It is tempting to connect the recent flu epidemic in Britain with a maximum or imminent maximum of solar activity. The rise of sunspot numbers towards this peak is possibly amongst the steepest on record. Although the new wave of flu and flu-like illness has not yet assumed pandemic proportions, the chances of this happening within a

single complete cycle of terrestrial seasons must be reckoned to be high.

In conclusion we note that electrical fields associated with intense solar winds can rapidly drive charged particles of the size of viruses down through the exposed upper atmosphere into the shelter of the lower atmosphere, the charging of such particles being due to the photoelectric effect. This could define one possible

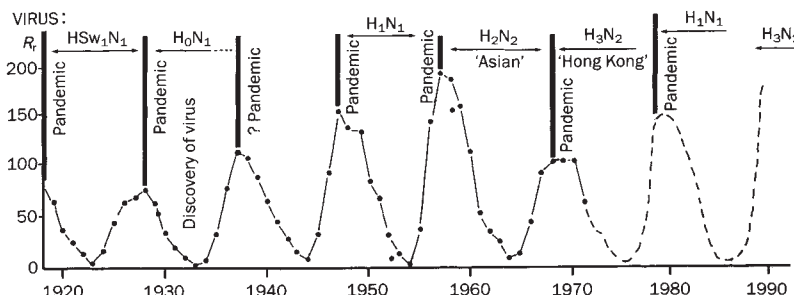
causal link between influenza pandemics and solar activity.

F. HOYLE

N. C. WICKRAMASINGHE

University of Wales,
School of Mathematics,
Senghenydd Road,
Cardiff CF2 4AG, UK

1. Hope-Simpson, R.E. *Nature* **275**, 86 (1978).
2. Beveridge, W.I.B. *Influenza, the Last Great Plague* (Heinemann, London, 1977).
3. Hoyle, F. & Wickramasinghe, N.C. *Diseases from Space* (Dent, London, 1979).



Yearly means of daily sunspot relative numbers compared with dates of influenza pandemics. The record up to 1971 is from Hope-Simpson; the dashed curve shows the situation for the period 1971–89.

Compound Q

SIR—Robert Buder in his article on the use of ‘Compound Q’ by AIDS patients in a programme organized by Project Inform (*Nature* **341**, 247; 1989) added that a formal clinical trial of Compound Q was organized in May. We wish to emphasize that the materials undergoing evaluation in Project Inform’s study and in a clinical trial approved by the Food and Drug Administration (FDA) are different, and that neither can accurately be described as Compound Q.

Trichosanthin is a protein isolated from the root tubers of *Trichosanthes kirilowii*, which has been used medically in China for a variety of clinical indications, chiefly induction of abortion. The name Compound Q originated as an informal, internal laboratory designation for trichosanthin during our initial collaborative *in vitro* studies of the material, at the University of California, San Francisco, and at Genelabs Incorporated, in Redwood City, California. Genelabs has developed a highly purified, formulated form of trichosanthin designated GLQ-223, now being produced by Genelabs according to FDA guidelines, which is the subject of an Investigational New Drug exemption granted by the FDA. Genelabs does not manufacture anything that can correctly be called Compound Q. There are no FDA-approved clinical studies of Compound Q currently being conducted at San Francisco General Hospital or anywhere else, nor are there any non-FDA approved studies being conducted with GLQ223.

The drug substance used by Project

Inform was a trichosanthin-containing preparation produced in the People’s Republic of China, where the drug is used as an abortifacient. In the interest of avoiding confusion, we believe this material should be described as such, not through the use of the vague and imprecise term Compound Q, which implies equivalence with GLQ223.

We believe that so far there are insufficient data to comment responsibly on either efficacy or toxicity of trichosanthin in HIV (human immunodeficiency virus) infected patients. As yet, neither the results from the FDA-approved trials of GLQ223 nor so far as we know, the results from Project Inform’s activities, have clearly documented either a therapeutic benefit of trichosanthin treatment or the lack of such a benefit. On the basis of anecdotal information, some HIV-infected patients may potentially be predisposed toward serious clinical events, including the development of coma, following administration of trichosanthin. We believe that until additional information about the possible risks and benefits of administration of trichosanthin to HIV-infected patients is obtained, such administration should occur only under closely controlled circumstances and under the supervision of a physician, according to a protocol approved by the FDA.

MICHAEL S. McGRATH

San Francisco General Hospital,
San Francisco, California 94110, USA

JEFFREY D. LIFSON

Genelabs Incorporated,
505 Penobscot Drive,
Redwood City, California 94063, USA

The peripatetic fossils: part 4

These two articles are the latest round in exchanges over the charge that Professor V.J. Gupta is responsible for corruption of the palaeontological literature on the Himalayas. Gupta (p. 307) responds to the allegations of four co-authors of papers with him. Professor J.B. Waterhouse, another co-author, comments below both on those allegations and on the articles by Dr John Talent in which the issue was raised originally. Talent will reply in next week's *Nature*.

A case of exaggeration

J.B. Waterhouse

JOHN Talent declares¹ that the scientific world has been deceived by a Houdini of the Himalayas, Professor V. J. Gupta. But is the story true? Certainly not in full.

One assertion may be readily discounted. Although Dr Talent seems mesmerized by the number of publications bearing V. J. Gupta's name, Gupta's contribution to geological literature is not anomalously huge and by no means approaches that of irreproachable Himalayan palaeontologists such as F.R.C. Reed or C. Diener (see table over). And when the large amount of excellent Chinese research in Tibet (Xizang) is taken into account, there is no way that V. J. Gupta can be said to have overwhelmed geological or palaeontological study on the Himalayas. The claim that the distribution of fossils has been irretrievably or disastrously deformed is itself an exaggeration, and if Gupta's work had put global distributions wildly out of kilter, then sound fossil distribution studies should reveal the falsification^{2,3}.

I can personally vouch that much of Gupta's material emphatically came from the Himalayas for parts at least of four periods of Earth history. He has sent me inarticulate brachiopods from the Tal Formation, purportedly Cambrian. They were accurately located, as identical fossils in similar rock are kept at the Geological Survey of India at Lucknow. For the Carboniferous, he sent me substantial collections purporting to be from the Fenestella Shales of Kashmir. I have compared the collections with type material from the same location^{13,14,17} housed at the Geological Survey of India, Calcutta, and British Museum (Natural History), London. The material was identical. I cannot allow that there could be any shred of doubt over their source. In our major joint study some 14 named species were described⁶, and of these 11 had already been reported from Kashmir¹³. Our two studies on the faunas clarified the age and generic relationships, but certainly did not add much to the diversity of fauna.

For the Permian, Professor Gupta has sent me about 40 different collections. Very few species were found to be new:

most have been identified, by me, with species from the Himalayas and nearby Salt Range (Pakistan), on the basis of direct comparison with my own large collections from the Himalayas in Nepal, and collections at the Geological Survey of India, Calcutta, British Museum (Natural History), London, and Nanjing Institute of Geology and Palaeontology. For example, from the Marbal Pass, Kashmir⁸, we described nine named species. Only one was new: the others were known elsewhere in the Himalayas and Salt Range. Gupta has sent me collections from Peninsular India (Umaria, Badhaura, Bap and so on). These collections are identical in lithology and fossil content with other well-sourced collections, so there is little doubt about their origin. One of the most interesting Permian faunas with some new species came from the Bijni tectonic unit of Garhwal Himalaya⁷, completely unknown to early workers such as Diener. It has received independent attention from other Indian palaeontologists¹⁸ and undoubtedly was correctly located by Gupta. Indeed Dr Talent himself has a very large collection of the same material, from the same general location: there can be no question about its source.

Finally, for the Triassic, an ammonoid has been reported⁹, again with confirmatory and independent evidence for source — the species, *Otoceras woodwardi*, is essentially limited to the Himalayan region.

In summary, for four periods, the general Himalayan source of Gupta-provided material has been verified independently by me. My own extensive field work in the Himalayas of Nepal, and to a minor degree elsewhere in India, and detailed distribution studies^{2,3}, help to confirm source and distribution. Dr Talent¹ has asserted that much of Gupta's palaeontological work is "based on forms never previously encountered in India or the rest of Asia", but this is certainly not true for the collections of Upper Palaeozoic faunas sent to me. Except for one athyrid (itself very close to an Iran-Armenian genus), I cannot think of a single genus amongst the species sent that

is not known elsewhere in Asia. Indeed about 90 per cent of them are known from elsewhere in the Himalayas. Overall, the Carboniferous-Permian picture is reasonable for regional accuracy, and in my view Professor Gupta has contributed positively to studies of Himalayan fossils. This is particularly welcome in view of the rarity of Indian palaeontologists prepared to publish good systematic studies. That is why the onus has been placed on non-Indian workers to help fill the gap, making a sad contrast with the Chinese contributions to the palaeontology of Tibet.

Not that I am satisfied with stratigraphic detail provided by Gupta. It is often too coarse and too rushed. Haste leads to the possible inversion of sequence, as several Indian geologists pointed out in one Zanskar sequence, for example, and I underlined in a report on *Eurydesma*⁴. Indian geologists are severely constrained by military strictures against publishing geological maps; even so, the lack of careful measurements and provision of only generalized stratigraphic and locality information are unsatisfactory. In several papers I have written with Gupta I have stressed the need for more detail. Professor Gupta is not alone in this failing, nor is India the sole nation at fault. Anyway, it is not the quality of his work that is being highlighted — such strictures would apply very widely. Rather it is its alleged fraudulent nature that is at issue.

What evidence has been offered that various experts should be deemed dupes? Not one of Gupta's fossils has actually been examined by Talent at first hand. Nor have any geochemical tests been carried out which might help discriminate between the alleged fossils from bazaars and real Himalayan material. There seems to be very little hard evidence. Not even counts are provided of the number of alleged fraudulent species. Dr Talent¹ makes frequent reference to a second article¹⁹ that allegedly substantiates the accusations, but this article also is written in flowery style, with unsubstantiated generalizations, and not a single detailed morphological and comparative analysis of a fossil, nor any counts, or distribution studies. The article may be good journalism, but it makes strange science.

Perhaps most revealing is Talent's failure to examine Himalayan collections at first hand, at Chandigarh, and at the

Approximate species counts for Himalayan studies

	Himalayas Waterhouse & Gupta ⁴⁻⁹	Himalayas Waterhouse ¹⁰⁻¹²	Himalayas Diener, von Krafft ^{13,14}	Salt Range Waagen, Reed ^{5,14-16}
Carboniferous macro-invertebrate fossil species	49	6	61	—
Permian macro-invertebrate fossil species	66	198+30 in prep.	70	668
Lower Triassic ammonoids	1	97 in prep.	84	166

The figures are based on the papers referred to and a number of other briefer works (not referenced).

Geological Survey of India at Calcutta. The 'case' against Gupta is remarkably rich in bold metaphors and unproven assertions, and somewhat thin in scientific analysis. The 'case' should have followed serious documented research. Instead it has preceded it. My dismay at the lack of detailed study is deepened on reading Dr Talent's criticisms of Gupta's Muth lithologies, as perhaps offering proof of Gupta's lack of familiarity with Himalayan rock¹⁹. Yet, in response, I have to suggest that Talent himself seems to have examined Muth rock with remarkable lack of attention to detail. Gupta and his colleagues have provided a better description of Muth lithologies.

It is also regrettable that Talent and colleagues¹⁹ characterized the Langu Group, which I named, as consisting of quartzite and dolomite: it consists almost entirely of dolomite, with thin shales and limestones¹¹. Moreover, Talent *et al.*'s study is replete with contentious observations, aimed, one supposes, at proving that Gupta has not actually worked in the Himalayas. The authors emphatically deny that Gupta could have found haematized ammonoids on the basis that haematite does not develop in frost-river rock in the Himalayas. Yet, in the 1989 season, I collected haematized brachiopods from the Late Permian Senja Formation near Mesokanto Pass, Nepal, at an altitude of nearly 17,000 feet. Haematized fossils can be formed widely in Upper Palaeozoic and Mesozoic rocks of the Himalayas. To further underline their unfamiliarity with Himalayan and Asian geology, Talent and colleagues asserted that, where they occur, Fusulinacea are always found in abundance. This is certainly not true — witness, for example, the rare samples in the Wargal Formation of the Salt Range, and in many locations in Thailand.

I do not vouch for Gupta's discoveries: I simply note that his alleged finds are being discredited for false reasons. It is beyond my expertise to comment on all the fossils sent by Gupta to various people. But from my knowledge the case against him has been exaggerated, and at least a number

of his joint articles hold their validity. If so, then Gupta, and his co-authors, have been falsely tainted.

I should here also like to comment on two of the articles published in *Nature* of 7 September 1989, in which the authors endeavoured to embroil me in the controversy. A. D. Ahluwalia says that I must "speak up honestly and boldly and address the grave charges point by point" (ref. 20), a demand I find completely fatuous because Ahluwalia bases his complaint on a "confusing report", unsourced, that I have been to a restricted area in the famous geological region of Spiti. As Bassi admits in the accompanying article²¹, there is no record of my name in Spiti checkpoints — I have never been to Spiti, and have never claimed to have been there, and this is verified in field notes and diaries deeded to the Australian Bicentenary Historic Records Search and in reports provided for the University of Queensland and the State Government of Queensland. The reason for trying to implicate me appears to be outrage that I wrote to Professor Gupta, to express a little sympathy and humanity in the storm of adversity that was engulfing him.

Although Dr Ahluwalia writes of my contributions to "so many vague and highly anomalous reports from the Himalayas", he cites only four. One is a note on the Tidong Valley²² to which I clearly contributed only a fossil list and age discussion in the middle of "p.346". The astute reader will note that V. J. Gupta is the principal author. I wrote little of the article. The fossil list is of species from the *Lamnimargus himalayensis* zone of the Himalayas, none exceptional, and the age discussion seems unobjectional to me. The very brevity of my contribution signals caution on my part regarding observations about stratigraphic detail and locality. The second paper is a worldwide, not just Himalayan, review of the bivalve genus *Eurydesma*, and discussion of its morphology, palaeoecology and species, including types kept at the Geological Survey of India at Calcutta. There is brief mention of *Eurydesma* from the Malung beds, correctly assessed as

pre-dating the Kuling shales despite Ahluwalia's claim. Anyone who thinks there is anything improper in this paper²³ should read it.

The third study²⁴ deals with a mystery species first described over 100 years ago¹⁷. Again, there is no implication whatsoever that I was involved in collecting the material, and in his article Bassi in no way explains what his problem is. The paper is a technical description, coupled with wide-ranging comparative analysis. This material has its only match with Davidson-described material, which is kept at the British Museum (Natural History) and was collected from the Himalayas; it shows that at least some material provided by Gupta did come from the Himalayas. The fourth article is one that I do not have, and Ahluwalia has possibly mistaken a publication date. If it concerns a mid-Permian ammonoid species, I again had no part in collecting the material. This ammonoid would indeed be rather rare in the Himalayas, and I would like to see more data on its source: that is why I cursorily identified it only to genus level. Allied specimens have been reported nearby in Tibet²⁵.

Like some others in this regrettable affair, Ahluwalia seems to be out of his depth with regard to the Late Palaeozoic of the Himalayas. He asserts that the Po Formation is Lower Middle Carboniferous. It is emphatically older — Lower Carboniferous, mostly Viséan, with some Serphukovian⁶. He also asserts that "*Eurydesma* has never been reported from anywhere below the level of the Gechang (Kuling) formation". This is not satisfactory either. The Gechang-Kuling coupling is inadequate, and ever since the early reports by Diener¹³ it has been recognized that the Kuling faunas comprise the *Lamnimargus himalayensis* faunal assemblage. *Eurydesma* is never found in such faunas but occurs in distinctly older beds. References to *Eurydesma* consistently support such a position, and show an Early Permian, not Upper Carboniferous age²⁶ (as claimed by Ahluwalia), in Tibet²⁷, Pakistan²⁸ and Peninsular India^{29,30}.

In a discussion on the Malung *Eurydesma*⁴, ignored by Ahluwalia, Gupta and I underlined the need for revision and pointed to the inadequacies of descriptions of geographical location and structural complexity. We also noted the failure of so many Indian colleagues to publish adequate systematic underpinning for their palaeontological and stratigraphic claims — a flaw by no means limited to geologists in India. This is one of the aspects I find so extraordinary about the concerted attack on Professor Gupta. I do not excuse Gupta, nor pretend that there is not a case to answer over duplicated plates, inadequate stratigraphy and absence of detailed mapping. But his accusers are also guilty

of poor locality data, poor age-dating, poor if any systematics, negligible mapping and failure to provide distribution statistics. Even Dr Talent, after many years in India, has not published any maps other than small sketches of road-side outcrops, which we geologists call "mud-maps" or less politely "crap-maps", and he has published no systematics to sustain his contentions.

I have tried to inject more organized data and stratigraphic integrity into the system. I have repeatedly expressed dis-

content in published work with V.J. Gupta, have repeatedly checked his collections against other type collections from the Himalayas, and have conducted my own mapping programme (in Nepal) to provide a standard against which any falsehoods, or inadequate work, may be tested. That, I believe, is the approach that will ultimately unravel the truth about the fossils of the Himalayas. □

J.B. Waterhouse is in the Department of Geology and Mineralogy, University of Queensland, St Lucia, Queensland, Australia 4067.

1. Talent, J.A. *Nature* **338**, 613–615 (1989).
2. Waterhouse, J.B. & Bonham-Carter, G.F. *Can. J. Earth Sci.* **12**, 1085–1146 (1975).
3. Shi Guang Rong & Waterhouse, J.B. *Internat. Brach. Congr.* 2, Abstr. 1990 Dunedin, NZ (in the press).
4. Gupta, V.J. & Waterhouse, J.B. *Indian Geol. Ass. Bull.* **11**, 141–152 (1978).
5. Waterhouse, J.B. & Gupta, V.J. *Recent Res. Geol.* **3**, 145–187 (1977).
6. Waterhouse, J.B. & Gupta, V.J. *Contrib. Himalayan Geol.* **1**, 108–146 (1978).
7. Waterhouse, J.B. & Gupta, V.J. *Recent Res. Geol.* **4**, 410–437 (1978).
8. Waterhouse, J.B. & Gupta, V.J. *Indian Geol. Ass. Bull.* **12**, 1–42 (1979).
9. Waterhouse, J.B. & Gupta, V.J. *Contrib. Himalayan Geol.* **3**, 219–224 (1985).
10. Waterhouse, J.B. *Jb. Geol. B.-A. Sond.* **12**, 5–99 (1966).
11. Waterhouse, J.B. *Palaeontographica Abt.* **A160**, 1–175 (1978).
12. Waterhouse, J.B. *Indian Geol. Ass. Bull.* **16**, 111–151 (1983).
13. Diener, C. *Palaeont. Ind. (N.S.) Geol. Surv. Ind.* **5**, 1–135 (1915).
14. Pascoe, E.H. *A Manual of the Geology of India and Burma* **2**, 3rd edn, Vol. II, 485–1343 (Government of India, Delhi, 1959).
15. Kummel, B. *Bull. Mus. comp. Zool.* **134**, 361–429 (1966).
16. Guex, J. *Eclog. geol. Helv.* **71**, 105–141 (1978).
17. Davidson, T. Q. *J. geol. Soc. Lond.* **22**, 29–45 (1866).
18. Bhatt, D.K. & Singh, M.P. *J. palaeont. Soc. India* **26**, 1–5 (1981).
19. Talent, J.A., Goel, R.K., Jain, A.K. & Pickett, J.W. *Cour. Forsch.-Inst. Senckenberg* **106**, 1–57 (1988).
20. Ahluwalia, A.D. *Nature* **341**, 13–15 (1989).
21. Bassi, V.K. *Nature* **341**, 15–16 (1989).
22. Gupta, V.J. & Waterhouse, J.B. *Recent Res. Geol.* **8**, 345–350 (1982).
23. Waterhouse, J.B. & Gupta, V.J. *Indian Geol. Ass. Bull.* **15**, 1–19 (1982).
24. Waterhouse, J.B. & Gupta, V.J. *Recent Res. Geol.* **8**, 389–396 (1982).
25. Sheng, H. *Mission Franco-Chinoise au Tibet*, 55–75 (CNRS, Paris, 1984).
26. Foster, C.B. & Waterhouse, J.B. *Aust. J. Earth Sci.* **35**, 135–157 (1988).
27. Liu Benpei & Gui Xincheng *Earth Sci. J. Wuhan Coll. Geol.* **1**, 79–92 (1983).
28. Reed, F.R.C. *Palaeont. Ind.* **23**, (1936).
29. Sahni, M.R. & Srivastava, J.B. *J. Palaeont. Soc. Ind.* **1**, 202–214 (1956).
30. Sahni, M.R. & Dutt, D.K. *Geol. Surv. India Rec.* **87**, 655–670 (1959).

Energy, New Delhi, in 1971. A preliminary note on the fauna from the dark limestone exposed south-east of Surichun La was published jointly by us⁹. I visited this area subsequently and made additional collections. My fieldwork in the Bara Lacha Ban and Surichun La areas was carried out on a number of occasions after 1975. The Lower Carboniferous conodont faunas from Ladakh, Lahul and Spiti (Losar) are identical and occur in abundance in all the regions. Good sections of the Lipak Formation are exposed in Bara Lacha Ban and Surichun La areas. Ahluwalia cannot comment on my visits to various parts of the Himalayas because he joined Panjab University only on 3 August 1982.

The Lower Carboniferous conodonts and other fossils recorded from different parts of the Himalayas came from the localities from where they have been reported. Ahluwalia himself handed over the conodont assemblages to Budurov, as he was at that time working at the Geological Survey of India. His contention of my having used some of the material given to Budurov is baseless.

The plate containing Figs 1–3 accompanying the paper on the Lower Carboniferous corals from Bara Lacha Ban¹⁰ was transposed with that of corals from the Luneak Formation exposed near Tanze, Ladakh. I would have detected this if proofs of the paper had been sent to me before publication, but of course take responsibility for the transposition.

The specimens of *Eurydesma cordatum* were collected from the Malung Shales of Ladakh by members of the Manali-Leh Expedition (1969)¹¹, which recorded the presence of Permian outcrops near Sarchu bridge. Ahluwalia's claim that *Eurydesma* has never been reported from anywhere below the level of the Gechang (Kuling) Formation is not correct. The stratigraphic position of the Kuling Formation — *sensu stricto* and not in the sense this term has been used by Srikantia *et al.*, 1977 (refs 12,13) — is well established and there is no record for the occurrence of *Eurydesma* from this formation. *Eurydesma cordatum* is restricted only to the Lower Permian in the Himalayas and the Kuling Formation is comparatively much younger in age (Panjabian to Dzulfian). Ahluwalia¹ has said that the sequence exposed around Sarchu is of Triassic age, but provides no palaeontological data in support of his interpretation.

The stratigraphic position of the lower and upper boundaries of the Ganmachidam Formation differs in different sections^{14,15}. The Ganmachidam Formation and the overlying fossiliferous sequence is well developed in the Savita nala section near Losar, contrary to Ahluwalia's claim. One has to climb upstream to observe this section. Ahluwalia accompanied four students to Spiti Valley in 1988

A response to the co-authors

Vishwa Jit Gupta

In *Nature* of 7 September there appeared four articles^{1–4}, each by a former co-author of mine, casting doubt on papers we had published together. Three of the writers (A. D. Ahluwalia, S. B. Bhatia and Udai K. Bassi) were the first authors in the respective papers under discussion. My observations on each of the articles follow. (The attention of readers is also drawn to my response⁵ to Talent's original article⁶.)

"Upper Palaeozoic of Lahul-Spiti" (A.D. Ahluwalia). The Devonian conodonts from Spiti discussed by Ahluwalia were collected by him when he was working in the Geological Survey of India. The specimens were macerated by him personally at the Geological Survey of India, Chandigarh. After processing the digested samples, Ahluwalia brought the residue to the Geology Department, Panjab University, and showed it to me and K. J. Budurov, a visiting scientist from Bulgaria.

A preliminary note was published in 1981 (ref. 7) with Ahluwalia as the first author. Subsequently Ahluwalia asked Budurov to examine the Spiti samples with a scanning electron microscope in Sophia, which led to another paper⁸. Later, on 24 April 1984, Ahluwalia wrote

to me from ETH Zurich. The relevant passage is as follows:

Now about Budurov. He has been most callous with our material and has reduced us to a state of incredibility taking away precious material and hatching it for three years with no results in sight still. I have all the Kinnaur stuff and we can go ahead on our own. I suggest let us drop or threaten to drop him out, for his priorities are somewhere else. You please write to him frankly on this issue and tell him that if he can't within 1984 then we would like to cut off our bandwagons from his. *The Spiti stuff I do not have more however, and he must return it [my italics].* I am writing to him but I have little hope of his writing back. So better you write please.

In his new version¹, Ahluwalia says: "I macerated afresh the part of the Muth dolomitic limestone sample that I had kept for further work". Yet in the letter cited above, he said that he had no sample from Spiti. Until now Ahluwalia has maintained silence and taken credit for our joint publications. Now he tells a different story. Ahluwalia has published 12 papers jointly with me — is he now going to withdraw his name from them?

The record of crinoids and conodonts from the Lower Carboniferous rocks exposed near Surichun La was based on the collections sent by Sh. Ravi Kaul, then Scientific Officer, Department of Atomic

and the entire field-trip (including travel from Chandigarh to Spiti and back) lasted 20 days (from 16 August to 4 September 1988). During this period he claims to have visited parts of Kinnaur, the Pin valley and other sections in Spiti (including sections exposed around Losar). Yet at least 8–10 days must have been spent in travel for both the inward and outward journeys. Ahluwalia simply could not have had much time for field work around Losar.

"Early Devonian Ostracodes" (S.B. Bhatia). Bhatia's statement that the specimens of Devonian ostracodes from Kaurig (Kurig) were handed over to him by me before his departure for Britain in 1972 is not correct. I had no interest in ostracodes in 1972 and as such the question of my processing the samples does not arise. In fact, the material from Kaurig and other regions of Lower Spiti was jointly collected and no collection was made by any individual member of the field party. As the paper concerned says (ref. 16):

The material for the present paper was collected by the authors during their 1966 expedition to the Lower Spiti valley. No detailed mapping of the area could be carried out as it lies very close to the India–China International Border. A generalized geological map can not be given for obvious reasons.

This paper was presented orally by Bhatia at the International Symposium on Himalayan Geology, organized by the Geological Survey of India, in 1976. The proceedings were published in 1982, which leaves a span of 16 years between collection of the material and publication of the paper. Bhatia did not disown the publication or the material until 1989 (23 years after the collection), when the present controversy started.

Bhatia has referred to the Devonian ostracode paper in several of his subsequent review articles, and in none of them has he questioned collection of the specimens.

"The Kinnaur region" (Udai K. Bassi). I visited this area in 1974 long before Bassi or his colleagues started fieldwork there. I visited Kinnaur again in 1975 with Professor I. C. Pande, then Director of the Centre of Advanced Study in Geology, Panjab University, to study the effect of the 'Kinnaur earthquake' which had hit the region that year. I subsequently visited the Kinnaur valley for field work on a number of occasions⁵.

Bassi¹⁷ (also in Radhakrishna¹⁸) suggests the presence of Carboniferous rocks at Kaurig and emphatically says that it is not possible to get Devonian fauna from this section. Unfortunately, so far, he has not

provided palaeontological or any other supporting evidence to substantiate this claim.

The upper part of the Muth succession is well exposed in the Yulang valley and has yielded well-preserved Devonian fauna. The papers referred to by Bassi give the stratigraphy of the Muth Quartzite and associated rock formation and I have not commented in any of my publications on the structural history of this region. The statement of Bassi that I changed the lithocolumn supplied by him for our paper on *Neogondolella regale*¹⁹ is not correct. The lithocolumn was finalized by Bassi and Ahluwalia before being handed over to me. This lithocolumn was submitted for publication unchanged.

Bassi's contention⁷ that the report of Carnian conodonts from Khimokul La cannot be true, as no rocks younger than Lower Triassic exist at this pass on Indo-Tibet border, contradicts a paper of which he is a co-author (ref. 17): "In addition to the Anisian conodonts and other microfossils, the upper units of the Triassic succession of the Tidong valley [have] also yielded representatives of *Paragondolella polygnathiformis* of Carnian age". I reassert that sequences of Carnian age and younger are exposed in the Khimokul La sector of Kinnaur region. Moreover, Bassi himself²⁰ has recently stated that no rocks younger than *Ladinic* occur in the Khimokul La section.

Bassi should consider withdrawing two other papers on Palaeozoic fossils of the Baspa valley^{21,22} (Kinnaur) in which, as acknowledged by the authors, the fossil identifications were done by me. The paper on the Permian fossils²³ from the Tidong valley in Kinnaur is also based on identifications made by me.

"Breakdown of trust" (Philippe Janvier). The paper on the Devonian fishes from Ladakh was finalized jointly by myself and Janvier and submitted for publication²⁴. At the time of finalization of the manu-

script Janvier proposed a new generic and specific name for the fish specimen. A few days afterwards, writing from Stockholm, Janvier asked me to delay the publication of the paper so he could note the similarity of the Ladakh specimens to specimens in the collection of Dr Mee Mang Chang in Stockholm. The revised manuscript, with the changed identification to "? Osteolepididae gen. et. sp. idet.", was resubmitted for publication. A few months after the publication of the paper I received a short note prepared by Janvier indicating close similarity of the Ladakh specimen with *Youngolepis praecursor* (Zhang and Yu). This note was also published under joint authorship²⁵. At no stage did Janvier indicate (even after seeing Chinese material in Stockholm) that the specimen had not come from Ladakh. The systematic descriptions, text figures and plates were prepared exclusively by Janvier. If Janvier had any doubt in his mind about the specimen why did he resubmit the revised manuscript for publication?

It was at Janvier's initiative that we collaborated in some joint publications after he visited me in my laboratory in Chandigarh in August 1977. He comments⁴ that scientists

... need to publish as many papers as possible in order to become famous. Many professional researchers in palaeontology do so, because they are required to do so. Some write virtually the same paper in different languages, others publish accounts of a particular fauna, group by group, sometimes fossil by fossil. In France, researchers are encouraged to do this.

This remark demolishes his own stand.

Finally I would like to bring it to the notice of my geoscientist colleagues that I have never carried out fieldwork nor visited the localities in the People's Republic of China from where Dr Mee Mang Chang made collections of the fish specimens described by her. □

Vishwa Jit Gupta is in the Centre of Advanced Study in Geology, Panjab University, Chandigarh-160014, India.

1. Ahluwalia, A.D. *Nature* **341**, 13–15 (1989).
2. Bhatia, S.B. *Nature* **341**, 15 (1989).
3. Bassi, U.K. *Nature* **341**, 15–16 (1989).
4. Janvier, P. *Nature* **341**, 16 (1989).
5. Gupta, V.J. *Nature* **341**, 11–12 (1989).
6. Talent, J. *Nature* **338**, 613–615 (1989).
7. Ahluwalia, A.D., Gupta, V.J., Budurov, K.J. & Kanwar, S.S. *Recent Res. Geol.* **8**, 281–283 (1983).
8. Ahluwalia, A.D., Gupta, V.J., Budurov, K.J. & Kanwar, S.S. *Riv. ital. Paleont. Stratigr.* **88**, 169–180 (1983).
9. Gupta, V.J. & Kaul, R. *Panjab Univ. Res. Bull.* **26**, (I–IV) 83–87 (1975).
10. Gupta, V.J. *Geol. Soc. India J.* **27**, 223–224 (1986).
11. Gupta, V.J., Bisaria, P.C., Chadha, D.K., Mahajan, G., Kumar, S., Kashyap, S.R., Kochhar, N. & Virdi, N.S. *Publ. Cent. Adv. Study Geol., Chandigarh* **7**, 77–84 (1970).
12. Gupta, V.J. & Waterhouse, J.B. *Indian Geol. Ass. Bull.* **11**, 141–142 (1978).
13. Srikantia, S.V., Bhargava, O.N. & Kapoor, H.M. *Geol. Soc. India J.* **19**, 73–78 (1977).
14. Bhardwaj, A. *Nature* **341**, 379 (1989).
15. Gupta, V.J. & Waterhouse, J.B. *Contrib. Himalayan Geol.* **4**, 63–66 (1989).
16. Bhatia, S.B., Jain, S.P. & Gupta, V.J. *Geol. Surv. India Misc. Pub.* **41**, 285–292 (1982).
17. Chopra, S., Bassi, U.K., Gupta, V.J., Budurov, K.J. & Ahluwalia, A.D. *Recent Res. Geol.* **8**, 284–287 (1981).
18. Radhakrishna, B. *P. Geol. Soc. India J.* **34**, 561–563 (1989).
19. Bassi, U.K., Gupta, V.J., Chopra, S., Budurov, K.J. & Ahluwalia, A.D. *Indian Geol. Ass. Bull.* **21**, 155 (1988).
20. Bassi, U.K. *Geol. Soc. India J.* **34**, 587–595 (1989).
21. Bassi, U.K. & Chopra, S. *Contrib. Himalayan Geol.* **1**, 220–224 (1978).
22. Bassi, U.K. & Chopra, S. *Recent Res. Geol.* **8**, 188–195 (1981).
23. Chopra, S., Gupta, V.J., Bassi, U.K. & Ahluwalia, A.D. *Publ. Cent. Adv. Study Geol., Chandigarh* **12**, 303–305 (1980).
24. Gupta, V.J. & Janvier, P. *Recent Res. Geol.* **8**, 68–80 (1980).
25. Gupta, V.J. & Janvier, P. *Indian Geol. Ass. Bull.* **14**, 81 (1981).

Anatomy of acetone photolysis

A remarkable series of measurements of the photodissociation of a simple molecule points to a field in which experimentalists will have the advantage for a long time to come.

HERE, one might think, is a simple problem: take a simple molecule, give it more energy than its binding-energy by means of an absorbed photon from a sufficiently energetic laser and then work out how the energy will be distributed between the fragments. On the face of things, the problem is but another version of a task in elementary mechanics that young people might be expected to solve: an object is impulsively disrupted, when what matters is that momentum is conserved. But real molecules, even simple molecules, are more complicated than that.

The immediate difficulty is that even the simplest molecular fragments are not billiard balls, but have internal degrees of freedom, rotation for example, whose excitation energy is small enough to compete effectively for a share of energy that would otherwise be shed by translation. (Nuclear physicists are often luckier, in that internal excitation of their reaction fragments can often be ignored.) It cannot be for nothing, after all, that J.C. Polanyi and his associates at the University of Toronto won a Nobel Prize in 1986 for their elegant, but emphatically empirical, techniques for identifying the excitation state of the products of molecular reactions between species whose internal motions are themselves well defined.

When will it be possible to predict these things? Not soon, it seems. Even the simpler problem of telling how quickly, and by what route, an impulse of energy will be redistributed among several internal degrees of freedom in a molecule is still up in the air, 30 years or so after people first appreciated that lasers could provide impulses of energy to predictable modes of internal motion, rotations, vibrations or even electronic excitations. Working out the excitation state of the products of molecular photolysis will require that problem to be superposed on that of following the time-course of the decomposition.

Almost any issue of the *Journal of Chemical Physics* (which tends to publish more than 15,000 pages of text each year) will provide a snapshot of where the problem stands, but rarely more elegantly than in a study of the photodissociation of acetone by a group from Cornell University (Trentelman, K.A. *et al.*, *J. chem. Phys.* **91**, 7,498; 1989), itself the culmination of a long series of measurements. Acetone (C_3H_6O) is the simplest ketone,

consisting simply of two methyl groups attached to the carbon atom of a carbonyl group or, alternatively, of acetic acid in which the ionizable OH group has been replaced by a methyl group. Hardly anything, it might be thought, could be simpler.

One complication is that there are two ways in which a sufficiently excited acetone molecule can dissociate, either into a methyl (CH_3) and an acetyl radical or into two methyl radicals and carbon monoxide. The higher the excitation energy, the more important the second reaction.

Trentelman and her colleagues have used ultraviolet pulses at 193 nm from an argon fluoride excimer laser to excite acetone molecules in a carrier of helium; the energy is enough to ensure that the three-fragment reaction predominates. Indeed, the excess energy to be redistributed between translational and internal motion is no less than 221 kJ per mole (of acetone), more than a third of the initial energy of the ultraviolet photon (the equivalent of 619 kJ per mole). By good luck, the ultraviolet energy is only just enough to excite acetone molecules to the energy minimum of a bound, but necessarily metastable, Rydberg state, which means that molecules embark on dissociation with virtually no vibrational energy or that they are as nearly stationary as quantum systems can be, what with all that zero-point energy.

There is a host of questions to ask about the system, not the least of which is whether dissociation takes place in one step or two. (With less excitation energy, the end products are methyl and acetyl radicals, if only briefly.) The answer is that the two carbon bonds are broken sequentially, but that the interval between these events is on the average so much shorter than the timescales of other processes that it does not matter.

The strongest evidence for that is that no less than 28 kJ per mole of energy ends up as rotational energy of the CO fragment, more than twice as much as the rotational energy of the two methyl radicals. The simple but convincing argument is that the instantaneous disintegration of an acetone molecule, which is symmetrical about the C—C axis, would give the CO fragments translational but not rotational energy, but that if one methyl radical has already left, the break-up of the remaining asymmetrical fragment is almost designed to kick the CO into a high rota-

tion state — the most commonly occupied state has 32 rotational energy quanta. The same argument leads to the conclusion that the methyl radicals will have less rotational energy — each of them is bonded to the central carbon in a direction passing through its centre of gravity.

In reality, the bulk of the energy released (124 kJ per mole) goes into translation, inferred from measured Doppler shifts of rotational lines in the fluorescence spectrum of CO and, for the methyl radicals, from elegant time-of-flight measurements. (The total energy accounted for is some 15 per cent less than that calculated from the binding energy of acetone, which is within the limits of the errors.) The direct implication is that the interval between the excitation of a molecule and its dissociation cannot be long enough for energy to be equipartitioned, or even unequally redistributed, between the internal degrees of freedom of the intact molecule. That leads to the conclusion that the dissociation is an impulsive process, and in passing explains why the methyl radicals have less vibrational energy than might be expected.

The route to dissociation can also be determined from the measurements. In unexcited acetone molecules, the three skeletal carbon atoms and the oxygen lie in a plane, as do the same atoms in the state reached by absorption of a 193 nm photon, but there is an intermediate unbound state with a non-planar configuration. What the Cornell group says is that the excited state is first converted to the intermediate unbound state before dissociating.

The unnerving feature of this intricate and technically breathtaking study is that it has taken so much effort to reach a well-rounded picture of what is, after all, a reasonably simple molecular reaction. Elsewhere, in the same issue of *J. chem. Phys.*, are essays in the direction of *ab initio* calculations of some of the properties of such real systems — there is an attempt to show how classical calculations of the coupling between different vibrational modes of benzene would be affected by zero-point energy and a somewhat formal translation of Feynman's path-integral calculus to the calculation of reaction rates. Taken together, they suggest this is a field in which the experimentalists will have the upper hand for some time to come.

John Maddox

The moorland owners' grouse

John R. Krebs and Robert M. May

THE red grouse *Lagopus lagopus scoticus* is more than a 600-gramme member of the family Galliformes. It plays a central, if somewhat self-sacrificial, role in the grouse-shooting enterprise in Britain; an average of 400,000 grouse are shot each year and they generate a gross income of the order of £10 million. The economic rewards from grouse shooting and its associated industries are by no means negligible for those living in many rural communities of northern England and Scotland, but for many outside these areas the conservation issues related to grouse-shooting are of even greater importance. There are roughly one million hectares of grouse-shooting land in Britain, primarily heather moorland and blanket bog; these habitats are rare and becoming rarer on a national and Western European scale and they support an unusual flora and fauna. The conservation issue can be stated simply: if grouse-shooting cannot be made to pay (this requires a summer grouse density of about 140 birds km^{-2} and a concomitant bag of around 35 birds km^{-2}) the main alternative form of land use for these upland areas is forestry. Silent, gloomy ranks of Sitka spruce replace open moorland and the characteristic plants, insects, birds and mammals disappear. Sadly for the cause of conservation, and for those who make their living from grouse shooting, stocks of grouse appear to be diminishing. This was the background to a recent workshop*, the aim of which was to assess current knowledge about the dynamics of red grouse populations and the implications of this knowledge for harvesting and management.

Scientists from two research groups were brought together to discuss their results with experts from Britain and North America. One group, based at Banchory near Aberdeen, is part of the Institute of Terrestrial Ecology (ITE) and has been studying red grouse populations in north-east Scotland since 1956. The other, from the Game Conservancy (GC), has been studying the red grouse in northern England since 1979 and in Scotland since 1985. The two groups broadly agree on the demographic facts about red grouse populations, but they offer different views on why grouse numbers change and hence on how to manage grouse populations in order to maintain or improve yields.

Three potentially independent demographic patterns were accepted by both teams as being unequivocal. First, since the beginning of the twentieth century

there has been a gradual long-term decline in the number of grouse shot both in England and Scotland. In Scotland, the decline was particularly marked in the period preceding and during the Second World War and again in the 1970s and 1980s (see figure). Second, different moors characteristically support different average densities of grouse; for example, heather moor in Scotland has a higher density of grouse than moorland covered by peat bog in Ireland. Third, and perhaps most intriguing, in many — but not all — moors, grouse show cyclic oscillations in numbers, with peaks occurring every four years (northern England) or every six years (Scotland). These cycles appear to be roughly synchronized over wide areas.

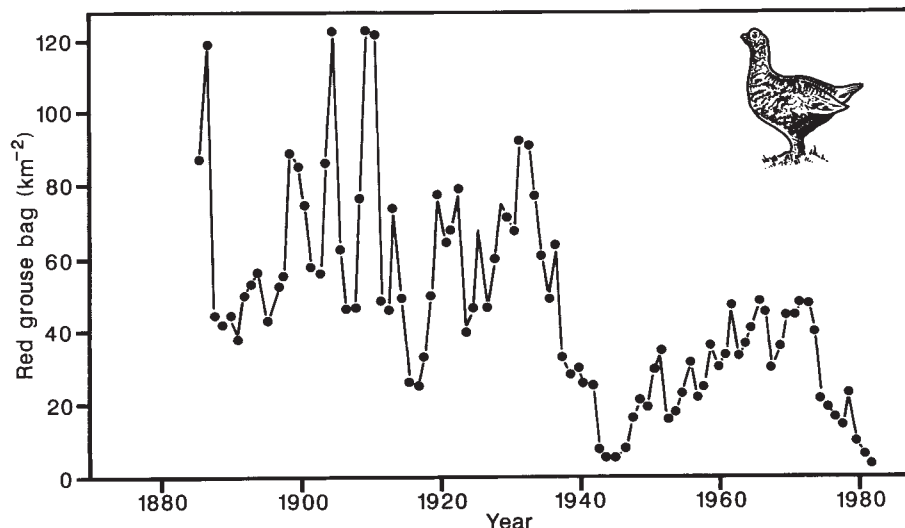
Both the ITE and GC groups agreed that the long-term decline in grouse is associated with habitat degradation. A combination of overgrazing by sheep and deer, improvement of uplands by drainage and the development of forestry has led to a decrease in the extent and quality of heather moorland, the grouse's preferred habitat and main source of food (young heather shoots). The effects of habitat degradation may be exacerbated by a concomitant increase in habitats for predators on grouse. Furthermore, management of both predator populations and extant heather moorland by gamekeepers is not as intensive as it was in the earlier part of this century.

Differences between moors in average population density may also be related, at least in part, to differences in vegetation structure. In general, moors that are less productive for grouse have little heather and often overlie acidic rocks, whereas

productive moors have plenty of heather and overlie basic rocks. Thus it might, to some extent, be possible both to halt the decline of grouse populations and increase the productivity of poor moors by replanting heather or encouraging its re-growth (government grants are now available for this) and by managing the heather carefully once it is established. Careful management in this context means burning small strips of the heather on a rotational basis so that the moorland consists of a fine-scale patchwork of young heather, which is nutritious and provides good food, and old heather which provides cover for the grouse.

The causes of cycles in grouse numbers are more enigmatic and here the two groups place emphasis on different causes. Four broad kinds of explanations were discussed at the meeting: behavioural changes associated with changes in average territory size, and interactions between red grouse and their parasites, their predators and their food supply. The essential feature of any hypothesis for cycles is that it must incorporate effects that exert 'delayed density-dependence' on grouse populations. This simply means that, as grouse numbers go up, the severity of the effects that depress recruitment or increase mortality must increase with a lag and then remain high long enough after the grouse have started to decline to sustain the decline for more than one season.

It is easy to understand why parasites might act in this manner. If their effect on grouse depends both on transmission rate (likely to be related to grouse density) and on the intensity of infection, then disease will become more common (with a slight lag) as grouse density increases and will remain high after it has started to cause decline in grouse numbers. This idea was first proposed by James Dalziel Dougal in 1875, in a report on a decline in grouse



Number of grouse shot per square kilometre in the Scottish Highlands over the past 100 years. Note the long-term decline; the relatively large declines around the two world wars; and the roughly cyclic changes within the longer term trends.

*British Ecological Society/Royal Society for the Protection of Birds workshop on red grouse population dynamics, Imperial College, Silwood Park, 13–14 December 1989. A summary will be published by the BES, and a fuller account by the RSPB.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Red grouse approach a butt on a rainy, misty day — if grouse-shooting cannot be made to pay, forestry may be the only alternative land use.

numbers during the 1860s and 1870s. The GC view is that a gut helminth, *Trichostrongylus tenuis*, causes fluctuations in grouse numbers primarily by reducing recruitment to the population through its effect on chick survival.

The evidence for this view is threefold. First, a ten-year study of Gunnerside Moor in North Yorkshire showed that both grouse and helminths show cyclic variations in their numbers. Second, populations to the west of the Pennines tend to be cyclic and to have high parasite loads, whereas those on the eastern side of the Pennines are less cyclic and have lower parasite loads. It is thought that this difference relates to rainfall: conditions on the wet western moors are more favourable for the survival of free-living stages of the parasite than are the drier conditions in the east. Third, experimental treatment of female grouse with an anti-helminthic drug increased their production of young and perhaps their survival (a result also obtained by the ITE group). It is not, however, straightforward to translate this evidence into the conclusion that helminths cause grouse cycles. The descriptive data do not disentangle cause and effect, and it is not yet clear whether the experimentally induced effects on chick survival are sufficient to drive the cycle. As a test of the hypothesis, P. J. Hudson has developed a method of coating grit, which is eaten by the grouse, with an anti-helminth drug and has persuaded two landowners in North Yorkshire to treat their entire moors with the coated grit. The cyclic decline in grouse numbers on the two treated moors was less marked than on untreated, nearby, controls, but it is still too early to say whether or not the treatment will actually prevent population cycles.

The ITE team, whilst acknowledging that parasites may affect production of grouse chicks, emphasizes the importance of behavioural changes in territorial grouse as a cause of cycles. By removing territorial cocks and associated hens, the team has shown that in many years (especially years of high density) territorial behaviour in the autumn determines the number

of breeders in the following spring. Many birds are evicted from the moorland by territory owners and these individuals often die or disappear from the local population. Crucial for translating this behavioural mechanism of limitation of breeding density into one that causes cycles is to propose a mechanism whereby territory size changes from one year to the next in a delayed density-dependent fashion. Here, the ITE team does not have any critical evidence, although it has some intriguing results from captive birds suggesting that relative aggressiveness changes from year to year and that the birds hatching from eggs laid in high-density populations are more tolerant than those hatching from eggs in low-density populations.

The ITE group had originally favoured the view that cycling within a moor may be related to changes in heather quality. But attempts to halt a cyclic decline in the 1970s by fertilizing heather failed, and the team has therefore rejected the idea of a plant-herbivore interaction as the cause of the cycle. But it was felt by the workshop that this hypothesis should not be ruled out altogether.

Finally, both groups agreed that predators such as foxes, crows and raptors are not likely to be the cause of cycles in grouse populations, because predators do not seem to be sufficiently abundant to cause grouse numbers to decline from the peak of the cycle. There was, however, some discussion of the possibilities that predators may have an effect on average grouse numbers and that they may keep them down when the population is already low for other reasons. Especially delicate, in this context, is the suggestion by the GC workers that predation by hen harriers may be a significant mortality factor for young grouse. If it were shown that hen harriers prevent grouse populations from increasing (and both groups emphatically agreed that it is premature to jump to this conclusion) conservationists would be presented with a 'Catch 22' situation — the conservation of heather moors and their unusual fauna and flora depends on increasing grouse

stocks; but the simple-minded way to increase grouse stocks would be to remove hen harriers, one of the species for which the moor is being conserved.

Even if we had a clear understanding of the dynamics of red grouse populations in general, and of the mechanism causing cycles in particular, there would still remain questions about how best to harvest grouse. Among other things, the answer depends on whether the aim is to maximize average yield or to reduce fluctuations in yields by eliminating cycles. These aims may conflict: for a population that is naturally cyclic, it can be shown that average yields are often maximized by a level of harvesting that leaves the population cycling (albeit with modified amplitude, period or both). Levels of harvesting that are high enough to suppress cycling may be beyond the so-called 'maximum sustained yield' (MSY) point, resulting in relatively steady yields but at a lower level. A more general issue is that harvesting will often affect a population's ability to withstand the effects of unpredictable environmental perturbations, so that variance in yield becomes relatively larger as the MSY point is approached. In short, various kinds of trade-offs must be weighed against each other in any harvesting plan.

Whatever the harvesting policy, there remains the question of moorland management. Because landowners already manage moors, it might be possible to advance our understanding by treating management as a deliberate experiment. It would be interesting, for example, to compare the outcomes of a factorial experiment in which different moors were either treated or not treated with anti-helminth drugs and where the heather was or was not subjected to traditional patchwork burning. Such management-as-experiment could both produce insights into fundamental ecological questions and help to secure the future of grouse moorlands. □

John R. Krebs and Robert M. May are in the Department of Zoology, University of Oxford, Oxford OX1 3PS, UK.

Effects of small volume melts

Martin A. Menzies

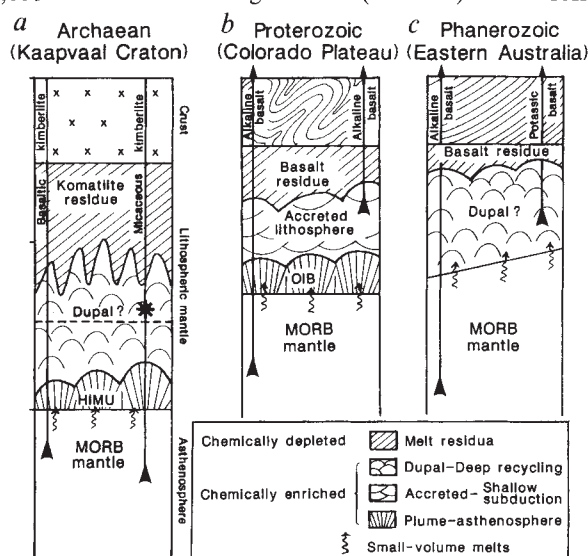
THE presence in the mantle of small-volume melt fractions (occupying less than 0.1 per cent of the mantle) was confirmed by the discovery of potassic, volatile-rich, micro-inclusions in diamonds and of other high-pressure inclusions. McKenzie¹ now shows in a theoretical study that these small-volume melt fractions are highly mobile and that, because of their propensity for concentrating incompatible elements, they can have a disproportionately large effect on lithospheric geochemistry and on magma compositions.

The clearest evidence of the effects of small-volume melts can be seen in the chemical stratification of continental lithosphere. Over the Earth's 4,000-million-year history, the extraction of melts and their residua from its interior has resulted in the formation of the rigid lithosphere, comprising crust (granite-basalt melt) and mantle (peridotite residue). This carapace has since been altered by the ingress of small volume melts from the hot asthenosphere (see figure).

The Kaapvaal craton of South Africa, dating from the Archaean (over 2.5×10^9 years old) and discussed by McKenzie¹, provides an excellent example. During its long exposure to the Earth's hot interior, metamorphism (alteration by heat and pressure) and metasomatism (chemical alteration by the influx of fluids or melts), have generated great isotopic and chemical heterogeneity (*a* in the figure). Micro-inclusions in diamonds², as well as other high-pressure xenoliths, indicate that material at 150–200 km depth has the isotopic characteristics of ocean-island basalts like those at St Helena. Mica, amphibole, rutile and ilmenite assemblages, originating from depths of 75–100 km, by contrast, have the characteristics of the 'Dupal anomaly' in the Southern Ocean³. This latter point may indicate the passage of the African Plate over the southern oceans sometime during the Proterozoic ($2.5-0.6 \times 10^9$ years ago) or Phanerozoic (up to the present). So widespread is the influx of small-volume melts that the uncontaminated protolith composition may survive only immediately beneath the crust-mantle interface between volcanic conduits. A further effect of the metasomatism is the presence of minerals throughout the Archaean craton whose apparent ages, determined through the isotope ratios of various incompatible elements, are Proterozoic and Phanerozoic⁴.

Whereas the sub-crustal Kaapvaal lithosphere, formed at high temperature through the extraction of komatiitic basalts to generate depleted peridotite, is 180–200 km thick⁵, the younger Proterozoic and Phanerozoic lithosphere, as found in the western United States and eastern Australia respectively, is chemically richer and less than 125 km thick. On the other hand, it has been exposed to the ingress of small-volume melts for only half the time. Nevertheless, both Proterozoic^{6,7} and Phanerozoic⁴ lithosphere seem to have become stratified.

Volcanic rocks and xenoliths are useful probes of the Proterozoic lithosphere of the Colorado Plateau. The protolith, a mid-ocean-ridge-basalt- (MORB-) like



Schematics to show effects of small-volume melt fractions. MORB, mid-ocean-ridge basalt; HIMU, high-uranium/lead ratio source material; OIB, ocean-island basalt.

peridotite, occurs immediately beneath the crust and is underlain by accreted lithosphere that attests to the more recent subduction and terrain-accretion processes. The garnet peridotites that form the deepest part of the lithosphere, however, are highly enriched in incompatible elements^{6,7} (*b* in the figure). The isotope composition here (at the 'thermal boundary layer') resembles that of the asthenosphere, presumably because of the action of small-volume melts.

The younger lithosphere of eastern Australia (no more than 600 million years old) has had a relatively uncomplicated history of chemical evolution (*c* in the figure). Small-volume melts from lower-mantle plumes similar to the Dupal plumes seem to be the sole factor. Compositions that are characteristic of the Dupal anomaly are superimposed on the depleted peridotite⁸ at the base of the original lithosphere. Several volcanic

rocks⁹ are thought to be derived from this layer, as are some basalt-borne xenoliths¹⁰. Alkaline volcanic rocks in Tasmania⁹ indicate that the asthenosphere in the region is MORB-like, so that the Dupal-type character of the eastern Australian lower lithosphere is a sure indicator of the dominant influence of small-volume melts from lower-mantle plumes.

McKenzie also discusses the extent to which the chemically stratified lithosphere acts as a primary or secondary reservoir for continental volcanic rocks. In a search for a repository for chemical-heat-producing elements, Oxburgh¹¹ suggested that alkaline basalts may originate in a hydrous lower lithosphere and tholeiites in the overlying shallow anhydrous lithosphere. The transfer of heat-producing elements to the lithosphere by small-volume melts produces discrete metasomatic horizons within otherwise refractory lithosphere. These metasomes

increase the 'fertility' of the lithosphere and act as a repository for the vital chemical fluxes required for magma production. Any collision between a metasomatized lithospheric plate and a high-temperature, upwelling, deep-mantle plume will result in partial melting of the metasomes and the eruption of highly potassic and alkaline magmas (lamproites and kimberlites) which may contain diamonds and high-pressure inclusions. Indeed, in the case of South Africa, the western United States and Scotland, these magmas can be used to map the spatial geometry of lithospheric domains in the mantle. Their regional distribution correlates to a significant degree with surface heat flow, crustal thickness and age province data⁷. Several examples of volcanic rocks of lithospheric origin exist: the Dupal horizon in the Archaean

keel beneath South Africa may have contributed to the genesis of micaceous kimberlites³ and lamproites; the accreted Dupal-like horizon under the Wyoming craton (older than 2.5×10^9 years) and the surrounding mobile belt of the Colorado Plateau may account for the origin of potassic and alkaline magmas^{7,12}; and the Dupal horizon beneath the Phanerozoic of Australia may have spawned leucitites⁹. In addition, in many of these terrains, alkaline (MORB) magmas are also erupted from the asthenosphere indicating that there are both shallow and deep sources for alkaline magmas in the mantle.

Flood basalts are the most voluminous volcanic rocks erupted on continents. *En route* to the surface they traverse Proterozoic and Phanerozoic lithospheres and the involvement of such lithosphere in their genesis is much debated. It is of interest to note that the potassium content

of these magmas¹ and their Sr, Nd and Pb isotopic parameters⁴ indicate the significant involvement of small-volume melts from the asthenosphere and possibly lower-mantle plumes containing a recycled crustal component, particularly in the case of the chemically enriched flood basalts associated with the break-up of Gondwanaland. It is becoming increasingly apparent that the huge volumes associated with flood basalts, require a major sublithospheric magma source and that the lithospheric mantle is not the most appropriate candidate.

Finally one can speculate that the chemical stratification in the Archaean, Proterozoic and Phanerozoic lithospheres records the tendency of continents to migrate from regions of upwelling of hot plumes and to congregate at regions of downwelling of cold slabs¹³. This would result in a vertical change in chemistry within the mechanical boundary layer from a shallow depleted protolith produced as a residue during intraplate volcanism (upwelling) to a deeper zone characterized by subduction zone chemis-

try (downwelling). This would be floored by a thermal boundary layer with the chemical characteristics of the asthenosphere (MORB) or upwelling lower-mantle plumes (see figure). □

Martin A. Menzies is in the Department of Geology, Royal Holloway and Bedford New College, Egham, Surrey TW20 0EX, UK.

1. McKenzie, D. *Earth planet. Sci. Lett.* **95**, 53–72 (1989).
2. Navon, O., Hutcheon, I.D., Rossman, G.R. & Wasserburg, G.J. *Nature* **335**, 784–789 (1988).
3. LeRoex, A.J. *Nature* **324**, 243–245 (1986).
4. Menzies, M.A. in *Continental Mantle* (ed. Menzies, M.A.) (Oxford University Press, in the press).
5. Boyd, F.R., Gurney, J.J. & Richardson, S.H. *Nature* **315**, 387–389 (1985).
6. Roden, M., Smith, D. & Murthy, V.R. *J. geophys. Res.* (in the press).
7. Menzies, M.A. *J. geophys. Res.* **94**, 7899–7915 (1989).
8. McDonough, W.E. & McCulloch, M.C. *Earth planet. Sci. Lett.* **86**, 327–340 (1987).
9. Ewart, A., Chappell, B.W. & Menzies, M.A. *J. Petrol. Special Lithosphere Vol.* (eds Menzies, M.A. & Cox, K.G.) 225–275 (1988).
10. Stolz, A.J. & Davies, G.R. *J. Petrol. Special Lithosphere Vol.* (eds Menzies, M.A. & Cox, K.G.) 303–330 (1988).
11. Oxburgh, E.R. *Geol. Mag.* **101**, 1–19 (1964).
12. Ormerod, D.S., Hawkesworth, C.J., Rogers, N., Leeman, W.P. & Menzies, M.A. *Nature* **333**, 349–353 (1988).
13. Anderson, D.L. in *Continental Mantle* (ed. Menzies, M.A.) (Oxford University Press, in the press).

GALACTIC STRUCTURE

Gas within gasless galaxies

Harley Thronson

THE gas and dust in space, the interstellar medium (ISM), is the stage on which the evolution of galaxies is played out: enriched by the products of dying stars, it is the material from which new generations of stars and planets are born. Sensitive telescopes working across the electromagnetic spectrum have made the study of the non-stellar matter one of the leading specialties in modern astronomy. A serious limitation has been an emphasis on a small minority of galaxy types, primarily giant spirals similar to our own Milky Way: these gas-rich systems are bright at many wavelengths, and their complicated gas dynamics and long history of chemical evolution present a formidable challenge. But a general understanding of the ISM in galaxies — its history and fate, its composition and abundance — will remain elusive until a much broader range of galaxy types is scrutinized. A persuasive example of the exciting future that awaits a more catholic investigation of the extragalactic ISM has just been presented by Young and Knezek¹, who report on the first large study of the relative abundance of the principal components, atomic and molecular hydrogen, of the cool ISM. The wide variation in composition between different galactic types may be hard to explain.

It had been widely believed for many years that non-spiral galaxies — that is, the majority of galaxies — possess no

ISM. Belief became a paradox in the 1970s as the fate of older stars became better understood. A simple calculation^{2,3} shows that over the period of time that has elapsed since the galaxies formed, commonly estimated to be 12,000–18,000 million years, as much 10^{10} solar masses of material lost by dying stars should have been returned to the ISM in giant galaxies of all types. This expectation is supported by observations of the gas-rich spiral galaxies. But 85 per cent of all the galaxies in the sky are not spirals and the majority of these are the ellipticals. More rare are the lenticular galaxies, possessing characteristics of both the ellipticals and the spirals. Both types of galaxies had been long regarded as entirely devoid of interstellar gas and dust.

In the past decade, astronomers have discovered 'gas-free' elliptical and lenticular galaxies, usually serendipitously, that in fact seem to possess massive amounts of gas. Presumably this is at least in part material lost by dying stars and, thus, the problem of galaxies without gas might be solved. Satellites launched to detect X-rays from space revealed⁴ that more-or-less normal elliptical and lenticular galaxies were often strong sources of this high-energy radiation, indicating large amounts of gas at temperatures of a few million degrees kelvin. Somewhat similarly, and equally surprisingly, analysis of data from the Infrared

Astronomical Satellite (IRAS) survey found that several hundred nearby elliptical and lenticular galaxies emit radiation characteristic of cool dust, probably mixed with at least moderate amounts of atomic or molecular gas^{5,6}. At the same time, detailed studies of the atomic hydrogen gas have been reported in 'gas-free' ellipticals and lenticulars^{7,8}.

Still one abundant — probably the most important — component of the ISM had yet to be found in these systems: molecular hydrogen, the gas out of which stars and planets are born. Frustratingly, there is no way known to detect H_2 directly in the cool, dense clouds anywhere in space. Molecular hydrogen possesses a pair of identical nuclei and, following the laws of quantum mechanics, does not emit radiation under the frigid conditions of the ISM. Unavoidably, then, astronomers study molecular clouds using the second most abundant molecule in space, carbon monoxide (CO), with a relatively-easily detected emission line at a wavelength of 2.6 mm. Finally, 15 years after CO had been detected for the first time in galaxies, three major surveys this past year reported on the CO emission from lenticulars^{9–11}. To derive the total molecular mass, astronomers often adopt relations found between the strength of the CO emission line and the mass of H_2 within well-studied interstellar gas clouds close to the Sun.

The conditions are now set for work to combine observations of the ISM in galaxies which cover a wide range of structure, as Young and Knezek attempt in their new paper. They derive the total molecular and atomic gas masses for about 150 objects and find that the ratio of molecular-to-atomic gas mass varies by up to a factor of 20 as a function of galaxy type, with galaxies once thought to be gas-poor or gas-free having, in fact, the largest fraction of their neutral gas in molecular form. These objects are found to have about four times as much molecular gas as they have atomic gas, compared with about equal amounts of both for systems like the Milky Way. A prime reason that many galaxy types were thought to be devoid of an ISM was that astronomers were, first, not observing these galaxies, and then for years looking for the wrong stuff. ►

1. Young, J.S. & Knezek, P.M. *Astrophys. J.* **347**, L55–L58 (1989).
2. Faber, S.M. & Gallagher, J.S. *Astrophys. J.* **204**, 365–378 (1976).
3. Mathews, W.G. *Astr. J.* **97**, 42–56 (1989).
4. Fabbiano, G.A. *Rev. Astr. Astrophys.* **27**, 87–138 (1989).
5. Bally, J. & Thronson, H.A. *Astr. J.* **97**, 69–78 (1989).
6. Knapp, G.R. *et al. Astrophys. J. Suppl.* **70**, 329–388 (1989).
7. van Driel, W. thesis, Univ. Groningen (1987).
8. van Gorkom, J.H. *et al. Astr. J.* **91**, 791–807 (1986).
9. Sage, L. & Wrobel, J. *Astrophys. J.* **344**, 204–209 (1989).
10. Thronson, H.A. *et al. Astrophys. J.* **344**, 747–762 (1989).
11. Wiklind, T. & Henkel, C. *Astr. Astrophys.* **225**, 1–11 (1989).

Why should different galaxy types show such wide variation in the relative amount of the two main components in the ISM? Young and Knezek debate a few possibilities: preferential destruction of H_2 relative to atomic gas and an enhanced concentration of gas near the centres of some types of galaxies. In the first case, exuberant star birth from molecular clouds taking place preferentially in some galaxy types should lead to the dissociation by starlight of H_2 into atomic hydrogen. Alternatively, lenticular galaxies

and their relatives possess massive spherical bulges of stars at their centres, which might gravitationally concentrate the interstellar gas, leading to the formation of molecules at the expense of the more diffuse atoms. More speculative possibilities include the roles of different dust compositions and large-scale gas dynamics as a function of galaxy type and location. □

Harley Thronson is currently visiting the Royal Observatory, Edinburgh EH9 3HJ, UK.

MOLECULAR BIOLOGY

Signals in the wounded plant

Dianna Bowles

STRESS or pathogen invasion evoke a number of defence responses in higher plants. Some responses are limited to the local site of stimulation, whereas others occur systemically, in regions of the plant far from the local stimulus. Many genes that are activated in these local and systemic responses have been characterized in detail, but far less is known of the signalling events that lead to their activation. Of particular interest is the sequence of events that link recognition of a stimulus by one cell to changes in gene expression in distant cells. A good example of this long-range signalling is the wound response of tomato and potato plants. Wounding induces the accumulation of novel gene products, including proteinase inhibitors (PIs) which are known to act as toxic deterrents to pathogens and pests, and which, when expressed constitutively in transgenic plants, confer insect resistance¹. The phenomenon, discovered by Ryan almost 20 years ago², has been extensively studied in the intervening years³. Now, in a paper published in *Proceedings of the National Academy of Sciences*⁴, Lothar Willmitzer's group puts an entirely new perspective on our understanding of the response.

To appreciate the significance of the new data, it is first necessary to understand the existing conceptual framework. This has been constructed from the following observations. First, wounding one leaf of a plant leads to activation of genes encoding PI proteins, both in the wounded leaf and in distant unwounded leaves. Second, removal of the wounded leaf by excision through the petiole prevents the activation of PI genes in the remainder of the plant. Because excision through the petiole is not sufficient by itself to trigger PI gene activation, a bioassay can be established in which compounds can be applied to the excised leaf and tested for their effects. Third, cell wall fragments, applied in the bioassay, induce PI gene activation. So evidence from wounding an intact plant shows that some-

thing moves out from the site of injury, and that the effects of this mobile signal can be prevented by detaching the source of the signal (the wounded leaf). Also, the bioassay indicates that oligosaccharides can mimic the effect of wounding as their application induces PI gene activation.

These two sets of observations have been linked in a long-standing model in which oligosaccharides represent the mobile signal. The suggested sequence of events therefore involves damage to a leaf-blade; fragmentation of structural cell wall polysaccharides in the injury zone; release of fragments (oligosaccharides) into the vascular system; transport of the oligosaccharides around the plant; and induction of PI gene activation in responsive cells throughout the organism. Despite a resounding lack of evidence for this model and some data to the contrary⁵, the role of oligosaccharides as the mobile signal has continued to attract attention. Certainly, application of exogenous oligosaccharides can lead to the activation of PI genes, and the signalling events involved are known to be conserved in a heterologous transgenic system⁶. Moreover both the wound response and the effects of oligosaccharides can be inhibited by aspirin and other agents that affect ion transport, suggesting that induction by either route shares at least one common event⁷.

Willmitzer's group⁴ introduced a different molecule into the regulation of the wound response: abscisic acid (ABA). Several lines of evidence are presented in the paper, three of which will be summarized here. First, spraying a leaf of an intact potato plant with ABA (10 μ M or 100 μ M) leads to accumulation of messenger RNA transcripts encoding PI protein II in the leaf that was sprayed and in unsprayed leaves; that is, direct application of ABA mimics the response induced by wounding. Importantly, the tissue specificity of the systemic response was maintained because neither wounding nor ABA led to PI gene activation in the root system and lower stem region of the plants. Second,

the response of ABA-deficient mutants was analysed. Both the potato (droopy) and tomato (*sit*) mutants have reduced levels of endogenous ABA, and spraying with the growth regulator reverses the mutant phenotypes. In both mutants, the levels of PI transcripts induced by wounding were very much lower than those induced in wild-type plants. But direct application of exogenous ABA to the mutants again led to high levels of PI transcripts. Third, wounding resulted in a substantial increase (doubling) in the level of extractable ABA, both in wounded leaves and in unwounded distant leaves of the wild-type plants. It is possible that the effect of the ABA spray is unrelated to events that occur *in vivo* on wounding, and represents only an alternative means of triggering the system that would not normally occur (for example, through the effect of ABA on transpiration, ion movement or both). But when the spray data are combined with the lack of responsiveness of ABA-deficient mutants to wounding, and the evidence that ABA levels increase locally and systemically during a wound response, a re-evaluation of the process becomes necessary.

Willmitzer and his colleagues are careful not to say that ABA is the mobile systemic signal, but they do point out that their data are entirely consistent with such an interpretation. In their model, ABA replaces the oligosaccharide signal. Injury leads to the release and increase in level of active ABA; transport from the local site occurs; the level of ABA in distant tissue increases; and the systemic response is induced. Direct spraying by ABA would be expected to mimic the wound, because ABA would increase locally, and as a consequence of diffusion into the vascular system, systemically. Transport of a mobile signal through the vascular system is compatible with data from wound-induced expression of hybrid genes in transgenic plants, in which the systemic activity of the PI promoter is initially confined to the vascular system (specifically, a channel of vascular parenchyma⁸). Because a systemic response *in vivo* is induced above and below the wounded leaf, movement must be bi-directional — that is, with and against the transpiration stream. Both roots and leaves can synthesize ABA, and changes in ABA content of xylem sap, at least, have been noted⁹. However, the vascular parenchyma is also the only known channel for conduction of electrical signals in plants, and the systemic induction of PIs by localized heat treatment has recently been correlated with long-range electrical signals¹⁰.

Interestingly, the effects of growth regulators on the induction of PIs have been examined previously. In an early study involving tomato plants¹¹, Ryan found that application of ABA had no effect on levels of PI protein; similarly,

Doherty¹² has been unable to show an increase in PI activity in response to ABA. It is possible that the age of the plants used and their growth conditions influence the effects of compounds applied, particularly because the wound response is known to depend upon nutrient status as well as temperature, light and levels of carbon dioxide¹³. But in a more recent study — this time on the expression of a chimaeric gene in transgenic tobacco callus, using a construct composed of a PI gene linked to a reporter gene — Kernan and Thornburg found¹⁴ that ABA (3.8×10^{-5} M) did not influence PI promoter activity. In contrast, ethylene (1.4×10^{-7} M) and cytokinin (BAP, 8.8×10^{-6} M) induced expression, whereas auxin (NAA, 2.0×10^{-6} M) actually inhibited the response. It may well be difficult to extrapolate data obtained with undifferentiated tissue to events that occur in intact plants, in particular to the possible role of these regulators in systemic events.

PLANETARY SATELLITES

An occult view of Titan

James Elliot

WHEN Voyager 2 bid farewell to the Saturn system in August 1981, many thought that its observations of Titan would be the last high-resolution data for that satellite's smoggy atmosphere until the next spacecraft visit — currently planned to be the Cassini mission which is scheduled to arrive after the turn of the millennium in 2002. After all, Saturn is so far from Earth that the best images of Titan hardly resolve the satellite as a distinct disk, owing to blurring by turbulence in the Earth's atmosphere. Even if these atmospheric effects could be overcome by interferometry or active optics techniques, a telescope with main mirror about a kilometre in diameter would be needed to give diffraction-limited resolution of Titan to rival the few-kilometre spatial resolution of Voyager's images. However, as reported on pages 350 and 353 of this issue^{1,2}, Titan's atmosphere has been probed with spatial resolution of only a few kilometres with the stellar occultation technique. With

But it seems that we are still some way from understanding the signalling events involved in this fascinating response. □

Dianna Bowles is in the Department of Biochemistry, University of Leeds, Leeds LS2 9JT, UK.

1. Hilder, V.A., Gatehouse, A.M.R., Sheerman, S.E., Barker, R.F. & Boulter, D. *Nature* **330**, 160–163 (1987).
2. Green, T.R. & Ryan, C.A. *Science* **175**, 776–777 (1972).
3. Ryan, C.A. & An, G. *Pl. Cell Environ.* **11**, 345–349 (1988).
4. Peña-Cortés, H., Sánchez-Serrano, J.J., Mertens, R., Willmitzer, L. & Prat, S. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9851–9855 (1989).
5. Baydoun, E.A. & Fry, S.C. *Planta* **165**, 269–276 (1985).
6. Sánchez-Serrano, J.J., Keil, M., O'Connor, A., Schell, J. & Willmitzer, L. *EMBO J.* **6**, 303–306 (1987).
7. Doherty, H.M., Selvendran, R. & Bowles, D.J. *Physiol. Mol. Pl. Path.* **33**, 377–384 (1988).
8. Keil, M., Sánchez-Serrano, J.J. & Willmitzer, L. *EMBO J.* **8**, 1323–1330 (1989).
9. Neales, T.F., Masia, A., Zhang, J. & Davies, W.J. *J. exp. Bot.* **40**, 1113–1120 (1989).
10. Wildon, D.C., Doherty, H.M., Eagles, G., Bowles, D.J. & Thain, J.F. *Ann. Bot.* **64**, 691–695 (1989).
11. Ryan, C.A. *Pl. Physiol.* **54**, 328 (1974).
12. Doherty, H. thesis, Univ. Leeds (1989).
13. Green, T.R. & Ryan, C.A. *Pl. Physiol.* **51**, 19–21 (1973).
14. Kernan, A. & Thornburg, R.W. *Pl. Physiol.* **91**, 73–78 (1989).

these observations, Sicardy *et al.*¹ and Hubbard *et al.*² determine the thermal structure and extinction of a several-hundred-kilometre region of Titan's upper atmosphere not probed by Voyager instruments; their first results indicate average mesospheric temperatures² of around 180 K and a global asymmetry of Titan's hazes¹.

In contrast with traditional telescopic imaging observations, which record light reflected from a body, stellar occultation observations record the starlight transmitted through the ring system and/or atmosphere of a planet or satellite. Because the spatial information in stellar occultation data is contained in the time history of the measured intensity of the starlight (and not from the image of the star or body formed by the telescope used for the observations), occulting bodies in the outer Solar System can be probed with a spatial resolution over a thousand times better than is possible by traditional imag-

ing methods from telescopes on Earth. The hitch, of course, in using the technique is that the body that one wants to study must be conveniently passing in front of a star that is bright enough to yield useful data.

The modern era of stellar occultation studies owes its beginning to Gordon Taylor, who, beginning in the early 1950s, compared planetary ephemerides with star catalogues to produce lists of bright stars that would be occulted by planets. The frequency of Taylor's predicted occultations was about one event every four years — hardly enough to qualify the occultation technique as a routine tool for planetary astronomers. In fact, on average, a star as bright as 28 Sagittarii, used in the new work, is occulted by Titan only once in 80 years.

The problem of finding useful stellar occultations and better ways to observe them has been steadily assaulted in several ways during the past decade and a half, and use of the technique has become more frequent and fruitful. Direct searches for occultation candidate stars on photographic plates and the observation of occultations in the infrared are significant improvements. Also, because the information that can be extracted from occultation data sometimes depends more on the location of the telescope than its size, and one can learn even more from multiple observations of the same event, much of the work is done from mobile observatories: small, portable telescopes (as was used for one station by Hubbard *et al.*), NASA's Kuiper Airborne Observatory and even Voyager itself. Some notable results from stellar occultations in recent years have been the discovery of the Uranian rings³ and the construction of a precise kinematic model for their orbits⁴; the discovery of Neptune's ring-arcs⁵ (since shown to be complete rings by Voyager); the systematic determination of the diameter and shapes of asteroids⁶; and the first direct detection of the atmosphere of Pluto^{7,8} — the only known planet not yet visited by spacecraft.

Shrouded by thick hazes, Titan's atmosphere — as we have learned from Voyager data — has two properties more similar to the Earth's atmosphere than any other Solar System body: the main constituent is

28 Sagittarii, on the left limb of Saturn, before (left) and during (right) the occultation by Titan, on 3 July 1989. The event, which lasted for just 6 minutes, was a rare chance to study the moon's 'ignosphere'. (The shading to the right of Saturn is an artefact.)

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

nitrogen and the surface pressure is 1.5 bar (compared with 1.0 bar for the Earth). Also like Earth, Titan probably has oceans⁹, but these contain ethane rather than water. The new measurements apply to an altitude range 250–500 km above the surface, a region called the 'ignosphere' by some, as only now has it been directly probed. The new results for the temperature and pressure in this zone will be used in the construction of more reliable atmospheric models for Titan, and these models will be crucial in the design of the Titan atmospheric probe that will be delivered by the Cassini spacecraft.

Occultations of stars by planets, satellites and rings are occurring as you read this, but probably none is being observed because the stars are too faint. However, with the increasing availability of more sensitive detectors and larger telescopes located at better sites — such as in Earth or planetary orbit — we can expect the outstanding results from stellar occultations to continue. The Hubble Space Telescope, soon to be launched, should facilitate the further observation of Neptune's tenuous rings, for example.

NASA's planned new Stratospheric Observatory for Infrared Astronomy (SOFIA) would allow systematic occultation studies of Pluto and Neptune's satellite Triton, as the aircraft can easily be manoeuvred to place the large telescope within the relatively narrow occultation shadow of these small bodies. Their meagre atmospheres undergo seasonal changes — as does Titan's — that we do not yet understand, and a steady flow of high-resolution atmospheric data are needed to help to solve these puzzles. Infrared arrays have created the opportunity for regular observations of occultations by Saturn's rings from Earth¹⁰. These will provide the data base needed to prepare for the Cassini mission by developing a comprehensive kinematic model for Saturn's rings, which has been possible to only a limited extent with Voyager observations, as these span only a narrow range of time. In planetary astronomy, the occult is becoming fashionable. □

James Elliot is in the Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Insight into blindness

Meredith L. Applebury

THE pace is remarkable; the molecular information is exact. Base by base, geneticists and molecular cloners are spelling out the specific molecular defects in the human genome that underlie inherited disease. Now, hard on the heels of the discoveries of the defects causing Duchenne muscular dystrophy and cystic fibrosis, comes the identification of a specific gene product and the molecular change that causes one form of retinal degeneration, a slow progressive dystrophy leading to blindness. On page 364 of this issue¹, Dryja and colleagues report that a point mutation in the rhodopsin gene leads to autosomal dominant retinitis pigmentosa (ADRP). The alteration is a base transversion of C to A, resulting in the change of a proline to a histidine residue at position 23 (see figure).

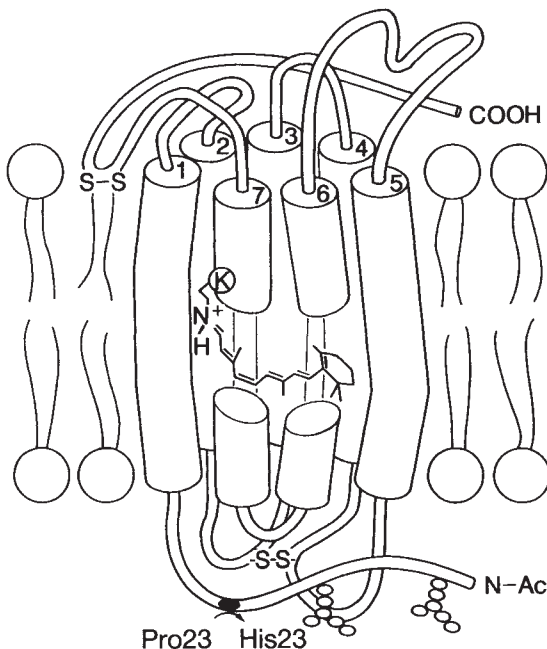
Retinitis pigmentosa is now known to be a collective set of inherited defects². Each defect ultimately causes distinct pigmented spicules to form across the surface of the retina, but this probably occurs long after the photoreceptors and

other cells have degenerated. Classification of the different types of degeneration is difficult, particularly for sporadic cases, but most of the disorders are thought to have underlying hereditary origins. Well-documented families show autosomal dominant, autosomal recessive and X-linked inheritance patterns. At a conservative estimate, one in every 4,000 people worldwide is affected with this relatively common eye disease. Retinal degenerations are one of the main causes of blindness in the human population.

The genetic defect often affects the retina or associated retinal pigment epithelium, but not other tissues or organs. For these cases, genes that are expressed only in the retina, or play important roles in retinal function, are suspects. Indeed, there are a host of candidate genes; rods and cones are specialized cells, with complex sets of proteins for detecting light and triggering physiological visual signals that can be processed and sent on to the brain. Many of the genes or complementary DNAs

encoding these molecules have been cloned, including those for the visual pigments that detect light and for the protein signalling components that they activate³. There are specific genes for rod and cone visual pigments, G proteins and cyclic GMP phosphodiesterases, all of which appear to be expressed only in photoreceptors. In addition, from biochemical work it seems that special forms of kinases, ion channels and sodium/calcium exchangers are involved in or regulate the signalling process, and that they differ from those in other tissues. In the past few years these candidates have been used as probes to seek genetic alterations in the genomes of individuals affected with retinitis pigmentosa⁴. Those tested so far have revealed nothing — at least no major deletions, rearrangements or telling polymorphisms.

The breakthrough came from Peter Humphries and his colleagues at Trinity College, Dublin, and their collaborator Stephen Daiger at the University of Texas, Houston⁵, who mapped the defect to the long arm (q region) of chromosome 3 in a large Irish pedigree with an autosomal dominant form of retinitis pigmentosa. Two



Rhodopsin serves as the molecular photoreceptor in rod cells of the retina. It is composed of an integral membrane protein opsin and the vitamin-A-derived chromophore 11-*cis* retinal, which is covalently attached to lysine residue 296. The opsin polypeptide chain spans the lipid bilayer seven times and folds in three dimensions to form a binding site for the chromophore. Upon absorption of light, rhodopsin is activated and triggers a cascade of reactions that decrease the second messenger cGMP concentration in the cell. The decrease elicits hyperpolarization of the cell's plasma membrane — the neurophysiological visual signal. A single base change in the opsin gene causes histidine to be substituted for proline at position 23. Over a period of time the photoreceptor cells degenerate, causing blindness.

1. Sicaudy, B. *et al.* *Nature* **343**, 350–353 (1990).
2. Hubbard, W.B. *et al.* *Nature* **343**, 353–355 (1990).
3. Elliot, J.L., Dunham, E.W. & Millis, R.L. *Sky Tel.* **53**, 412–416, 430 (1977).
4. French, R.G. *et al.* *Icarus* **73**, 349–378 (1988).
5. Hubbard, W.B. *et al.* *Nature* **319**, 636–640 (1986).
6. Millis, R.L. & Dunham, D.W. in *Asteroids II* (eds Binzel, R.P., Gehrels, T. & Matthews, M.S.) 148–170 (University of Arizona Press, Tucson, 1989).
7. Hubbard, W.B., Hunt, D.M., Dieters, S.W., Hill, K.M. & Watson, R.D. *Nature* **336**, 452–454 (1988).
8. Elliot, J.L. *et al.* *Icarus* **77**, 148–170 (1989).
9. Lunine, J.I., Stevenson, D.J. & Yung, Y.L. *Science* **222**, 1229–1230 (1983).
10. Harrington, J., Dunham, E.W., Forrest, W.J. & Pipher, J.L. *Bull. Am. astr. Soc.* **21**, 954 (1989).

photoreceptor genes became immediate candidates. The gene encoding the rod G-protein α -subunit maps to chromosome 3 (ref. 6). The specific location is unknown, but the syntenic relationship to the mouse chromosome location indicates that it might be in the short arm of 3. The gene encoding rod opsin maps to 3q (ref. 7). Dryja *et al.* took up the gauntlet and in just a few months they located the defect by searching for alterations in the opsin gene among their own well-documented ADRP patients, in particular among a pedigree of British descent.

Pairs of nucleotide oligomers were synthesized as primers enabling each of the exons in the opsin gene to be amplified and rapidly sequenced. The speed with which this can be done is made possible by using the polymerase chain reaction (PCR) and a few hundred nanogrammes of genomic DNA from a few drops of blood. Among the first 20 patients examined, five had one opsin allele with a C to A transversion altering the amino acid encoded at position 23 to a histidine. Homing in on this observation, Dryja's group synthesized two oligomers corresponding to the normal and the altered sequence. The allele-specific probes were labelled and hybridized with the genomic DNA to show the presence of the normal or the altered sequence. In 148 unrelated, affected ADRP patients, 17 carried the C to A transversion; in 102 random, unaffected individuals, all sequences were normal. Large families were examined more systematically; one, of British ancestry, showed cosegregation of the point mutation with the disease. The fact that other families with ADRP do not carry this polymorphism supports the contention that defects in other genes (or other alterations in opsin) lead to the same end point — retinal degeneration.

Proline 23 is conserved in all vertebrate opsins, most invertebrate opsins, and often in other related signal receptors such as the β -adrenergic receptor and some muscarinic receptors⁸. It would therefore seem that the proline is somehow functionally important. Proline frequently induces a kink or bend in a polypeptide chain, but it is unclear what role it plays in opsin and why a change to another residue leads to photoreceptor degeneration. The change might alter the folding properties of opsin, thereby affecting its function to detect light, catalyse the activation of G protein, or both, but there is no evidence to support this contention. The N-terminals of visual pigments from different photoreceptor cells and from different species vary widely; rhodopsins expressed artificially in other cell systems are glycosylated at nearby residues (positions 15 and 2) with large bulky complex carbohydrates, yet there is little effect on chromophore incorporation or activation function^{9,10}.

Other possibilities include effects on biosynthesis or degradation of opsin, ultimately causing excess wear and tear on the cell. The N-terminal must be transported across the membrane during biosynthesis, and a change in protein folding might slow the process or disrupt the addition of carbohydrate and its subsequent processing. Alternatively, proteolytic degradation of opsin that takes place in the adjacent retinal pigment epithelial cells may be adversely affected. The epithelial cells serve as caretakers for photoreceptor protein turnover; their malfunction would ultimately lead to photoreceptor degeneration. All this points to how little we know about a well-known molecule. Some insight may come by artificially expressing the mutant protein in cultured cells and creating appropriate animal models with the specific defect. Treatment or gene repair are still out of the question, and certainly will not be possible without a better understanding of the cell biology of the retina.

With rigorous linkage studies and a

host of candidate genes, the molecular defects leading to other forms of retinal degeneration are likely to be revealed soon. Meantime, one form of ADRP can be diagnosed specifically, which means that those who have a 50 per cent chance of passing on the defect to their children can now be offered accurate genetic counselling. □

Meredith L. Applebury is in the Eye Research Laboratories, University of Chicago, 939 East 57th Street, Chicago, Illinois 60637, USA.

1. Dryja, T.P. *et al.* *Nature* **343**, 364–366 (1990).
2. Heckenlively, I.R. *Retinitis Pigmentosa* (Lippincott, Philadelphia, 1988).
3. Falk, J. & Applebury, M.L. in *Progress in Retinal Research*, Vol. 7 (eds Osborne, N.N. & Chader, G.J.) 79–112 (Pergamon, Oxford, 1987).
4. Ringens *et al.* *Invest. Ophthalmol. Vis. Sci.* **30**, 43 (1989).
5. McWilliam, P. *et al.* *Genomics* **5**, 619–622 (1989).
6. Blatt, C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 7642–7646 (1988).
7. Nathans, J. *et al.* *Science* **232**, 203–232 (1986).
8. Applebury, M.L. & Hargrave, P.A. *Vision Res.* **26**, 1881–1895 (1986).
9. Oprian, D.D. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 8874–8878 (1987).
10. Khorana, H.G. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 7917–7921 (1988).

ASTRONOMY

Pulsars miss a beat

R.N. Manchester

THE outstanding characteristic of pulsars is undoubtedly the great stability of the basic pulsational period. Indeed, measurements of the period of the millisecond pulsar PSR1937+21 are limited by the stability of terrestrial clocks¹, raising the possibility that pulsars could be used to define the long-term standard of terrestrial time. Nevertheless, in younger pulsars especially, the period occasionally drops suddenly and without warning. But until the discovery of frequent, large 'glitches' in the period of PSR1737–30, reported by J. McKenna and A.G. Lyne of Jodrell Bank on page 349 of this issue², this phenomenon was hard to study because of its rarity and unpredictability.

The characteristic stability of these spinning neutron stars was exploited in the paper first announcing their discovery³: the constancy of the period in the heliocentric reference frame was used to demonstrate that pulsars must be outside the Solar System. The apparent period varies for pulsars orbiting in binary pairs, but the intrinsic period is very stable. In the case of the binary pulsar PSR1913+16, the stability made possible the first accurate determination of the mass of a neutron star (around 1.4 solar masses) and the first observational evidence of gravitational radiation⁴.

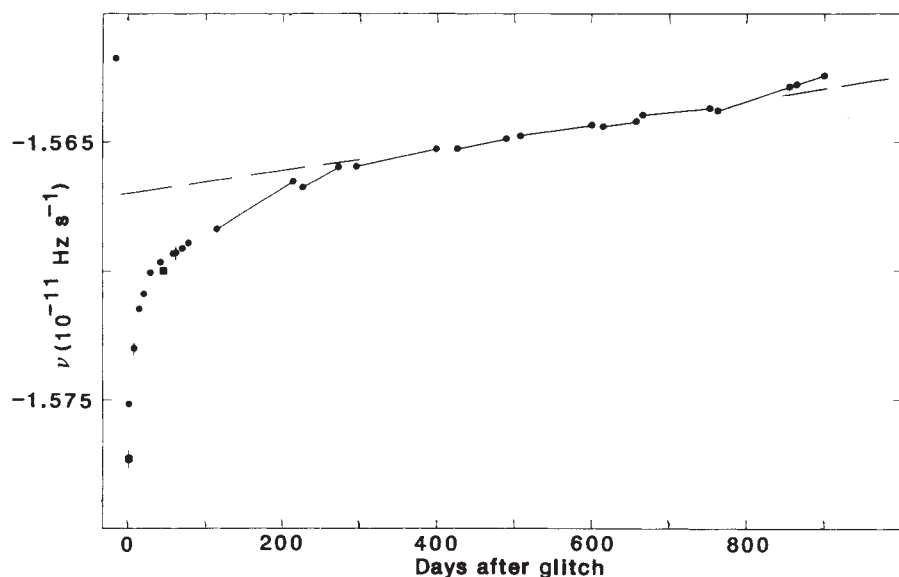
The Vela pulsar was the first to be observed^{5,6} suddenly to spin up and a further seven glitches have been observed in 20 years of timing of this object. Several other glitches have been observed in other

longer-period pulsars, but they are usually not recognized until long after the event.

The properties of glitches seem to depend strongly on the age of the pulsar. Those observed in PSR1737–30, Vela and other older pulsars are relatively large, with a fractional period change of the order of one part in a million. The period changes suddenly, within minutes.

All pulsars slow down very gradually, in most cases quite steadily. After a glitch, the period derivative is larger than the pre-glitch value so that the period tends back towards the extrapolation of the pre-glitch trend (see figure). For the Vela pulsar, about half of the initial period-derivative increment decays away in a few days⁷. Relaxation of the remainder of the increment is much slower, with a timescale of the order of one hundred days and is often interrupted by the next glitch⁸. Another pulsar which has experienced a large glitch (in fact, the largest ever observed), PSR0355+54, had a rather different post-glitch behaviour⁹. Almost all of the period-derivative increment decayed away with a timescale of about 45 days, leaving what may be a permanent step in the period.

In contrast, the four glitches observed in the younger Crab pulsar are much smaller, about a part in 10⁸ or less, and the increment in period derivative is similarly smaller. The decay of the period derivative increment appears to be a simple exponential with a timescale of about ten days. However, at least in the case of the



Recovery of pulsar spin-down (for PSR0833–45) after a glitch. The pre-jump frequency derivative is given by the point in the top-left corner and the two phases of recovery can be seen (from ref. 8).

1975 glitch¹⁰, not all of the period derivative increment decayed away, so that after about 50 days the pulsar was rotating *less* rapidly than it would have been had the glitch not occurred. Although the data spans are much shorter, no glitches at all have been observed in the two other known pulsars (PSR0540–69 and PSR1509–58) which are younger than the Vela pulsar.

Following the first Vela glitch, Baym *et al.*¹¹ suggested a model in which the glitch was due to a sudden cracking of the solid crust of the neutron star — a starquake — with a resulting reduction in the moment of inertia and hence spin-up of the pulsar. The suggestion that the interior of a neutron star is superfluid (made well before the discovery of pulsars¹²) and hence only weakly coupled to the crust, provided a neat explanation for the long timescales of the post-glitch recovery. This model was very successful in that it accounted for the increase in period derivative at the time of the glitch and its gradual decay. However, it could not cope with the recurrence of large Vela glitches every few years.

The model that currently seems best able to explain these large and frequent glitches involves pinning and unpinning of the normal-fluid cores of vortices in the neutron superfluid from crustal nuclei. A sudden unpinning of a small fraction of the cores and the consequent transfer of angular momentum to the crust accounts for the glitch and a gradual return of the creep rate of these cores to the equilibrium value accounts for the slow relaxation. By invoking two different regions of unpinning with different relaxation times, this model can account for the broad characteristics of the Vela glitches and the short and long recovery timescales following the glitch¹³. The different glitch characteristics of the very young pulsars

may result from higher temperatures in their interior allowing a smoother transfer of angular momentum from the superfluid core to the crust.

Although the vortex-creep model gives a satisfactory fit to much of the observational data, there is much that is difficult to explain. For example, the difference in the post-glitch behaviour of different pulsars does not have a natural explanation in terms of the model. In the case of the Crab, the persistent change in period derivative seems to imply a change in the processes of rotational energy loss from the star. It may be that some entirely different mechanism is responsible for the glitches in very young pulsars. In any case it is clear that careful study of the frequent glitches observed by McKenna and Lyne in PSR1737–30 will put new constraints on glitch models and lead to a greater understanding of the processes occurring in the superfluid interior of neutron stars. □

R.N. Manchester is at the Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 2121, Australia.

1. Rawley, L.A., Taylor, J.H., Davis, M.M. & Allan, D.W., *Science* **238**, 761–765 (1988).
2. McKenna, J. & Lyne, A.G. *Nature* **343**, 349–350 (1990).
3. Hewish, A., Bell, S.J., Pilkington, J.D.H., Scott, P.F. & Collins, R.A. *Nature* **217**, 709–713 (1968).
4. Taylor, J.H. & Weisberg, J.M. *Astrophys. J.* **345**, 434–450 (1989).
5. Radhakrishnan, V. & Manchester, R.N. *Nature* **222**, 228–229 (1969).
6. Reichley, P.E. & Downs, G.S. *Nature* **222**, 229–230 (1969).
7. McCulloch, P.M., Hamilton, P.A., Royle, G.W.R. & Manchester, R.N. *Nature* **302**, 319–320 (1983).
8. Cordes, J.M., Downs, G.S. & Krause-Polstorff, J. *Astrophys. J.* **330**, 847–869 (1988).
9. Lyne, A.G. *Nature* **326**, 569–571 (1987).
10. Demianski, M. & Proszynski, M. *Mon. Not. R. astr. Soc.* **202**, 437–451 (1983).
11. Baym, G., Pethick, C., Pines, D. & Ruderman, M. *Nature* **224**, 872–874 (1969).
12. Migdal, A.B. *Nucl. Phys.* **13**, 655–674 (1959).
13. Alpar, M.A., Anderson, P.W., Pines, D. & Shaham, J. *Astrophys. J.* **278**, 791–805 (1984).

Heavenly visions

LAST week Daedalus devised a process for making metallic snow, by cooling a metal vapour diluted with inert gas. The more dilute the vapour, the smaller the resulting solid particles. At a high enough dilution, a fine metallic 'smoke' results. Each particle is a single unbranched needle-crystal of metal, less than a micrometre long.

Daedalus sees his metal-needle smoke as the optical equivalent of the airborne strips of metal foil used to confuse military radar. Each strip is cut to a quarter wavelength of the radar transmission, and resonantly re-radiates it as an enormously enhanced echo. Similarly, a suspension of Daedalus's micro-fine metallic smoke will selectively scatter light for which its needles are quarter-wave resonators. Other wavelengths of light will be quite unaffected.

Released into the atmosphere, the new smoke would have intriguing effects. In cloudy weather, a release of blue-scattering smoke would selectively transmit yellow light, cheering and brightening the day. The blue light thus scattered would tinge the visible scene, 'expanding' the landscape with a false impression of great distance. In cloudless conditions, an ultraviolet-scattering smoke could guard sunbathers from sunburn, while an infrared scattering one could shield them from heat stroke.

But Daedalus despises such free, universal, compulsory socialist benefits. He wants his metallic smoke to make money. So DREADCO's physicists are releasing into the atmosphere thin, almost invisible suspensions of a blue-scattering smoke, and then 'writing' on them with a distant polarized infrared laser. Where it hits the smoke plume, the infrared beam aligns the needle crystals along its polarization axis. These regions of alignment become polarizers themselves. So they preferentially scatter the blue light of a cloudless sky, which is itself polarized, and show up as dark writing against the celestial blue. The image fades almost instantly as the aligned particles lose their orientation by random brownian decay. To maintain a still or moving picture, Daedalus will have to project a continuously scanning video-modulated infrared 'raster' on the invisible smoke plume, like the electron beam of a television receiver.

When perfected, DREADCO's polarized-laser sky-writing system will turn the whole wide blue sky into a TV screen! Advertisers should pay copiously to write slogans, images, or even complete commercials on the heavens. Holiday resorts would love to woo the electronic generation with the idyllic double pleasure of lazing on a sunny beach while watching endless TV-in-the-sky. Only the sound track, relayed from municipal loudspeakers, would remain earthbound.

David Jones

EBV Ig-like domains

SIR—Two proteins encoded in the genome of Epstein-Barr virus (EBV) possess regions with sequence similarity to the immunoglobulin domain of the immunoglobulin superfamily¹. We now report a third domain in p140, an EBV envelope protein.

A neural-network computer search program trained to recognize immunoglobulin domains² was used to scan

serologically and epidemiologically associated with exposure to several viruses, EBV⁴ included. One of us has previously proposed that non-specific 'mitogenic' activation of central nervous system T cells could result in bystander demyelination, and has demonstrated that activating anti-T-cell receptor antibodies are efficiently presented by glial cells *in vitro*^{5,6}.

It is possible that activating anti-T-cell

linked with MS⁷ has significantly similarity (s.d. > 3) with the carboxyl-terminal region of the EBV p140 immunoglobulin-like domain. It will be interesting to compare T-cell receptor amino-acid sequences of T-cell clones over-represented in the central nervous system or T-cell receptor susceptibility genes in MS, with protein sequences of other viruses associated with MS.

NEIL R. CASHMAN

Department of Neurology and
Neurosurgery,
Montreal Neurological Institute,
3801 University, Montreal,
Quebec H3A 2B4
Canada

YANNICK POULIOT

Department of Biology,
McGill University,
Stewart Biology Building,
1205 Avenue de Dr Penfield,
Montreal, Quebec H3A 1B1
Canada

1. Wang, H. Wu, J.-J. & Tang, P., *Nature* **337**, 514 (1989).
2. Bengio, Y., Pouliot, Y., Bengio, S. & Agin, P. in *Advances in Neural Information Processing Systems* (ed. Touretzky, D. S.) (Morgan Kaufmann, San Mateo, in the press).
3. Williams, A.F. & Barclay, A.N., *Ann. Rev. Immun.* **6**, 381–405 (1988).
4. Operskalskii, E.A., Visscher, B.R., Malmgren, M. & Detels, R. *Neurology* **39**, 825–829 (1989).
5. Cashman, N.R. & Noronha, A. *J. Immun.* **136**, 4460–4463 (1986).
6. Cashman, N., Boulet, S., Cragg, L., Bambridge, L. & Antel, J. *Ann N.Y. Acad. Sci.* **540**, 498–500 (1988).
7. Seibon, E. *et al. Cell* **57**, 1095–1100 (1989).

QQBE1 EV L Q T L R S M N Y PAVLG R V G E Q G - P D Q M F E V Q H G P E T V L R Q A L R L L G T W S S F A S E Q Y E C
A24574 EV I Q F L V S I S - Y D G T V - R K - E S G I P S G K F E V D R I P E T S T S - T L T I H N V E K Q D I A T - - Y Y C

Local alignment of the carboxyl-terminal region of the immunoglobulin-like domain of Epstein-Barr virus envelope protein p140 (QQBE1) with a human T-cell receptor γ chain (A24574) (s.d. = 5.35 by ALIGN program, bias = 6, gap penalty = 6). Identical residues (34 per cent) are boxed; conservative substitutions according to Dayhoff Matrix (19 per cent) are denoted by bars.

amino-acid sequences stored in the National Biomedical Research Foundation (NBRF) protein database (release number 19). As well as detecting virtually all known members of the immunoglobulin superfamily, the program identified viral proteins with previously unrecognized superfamily similarity including the p140 envelope protein of EBV (NBRF filename QQBE1).

The immunoglobulin-like domain of p140 comprises 106 amino acids between the cysteines at positions 888 and 994, and is thus similar in length to some T-cell receptor variable (V) immunoglobulin domains³. Chou-Fasman and Garnier-Osguthorpe-Robson algorithms predict 5 β strands of 5–12 amino acids between the cysteines and 2 flanking β -strands, consistent with the secondary structure of an immunoglobulin fold⁴. Alignments between members of the immunoglobulin superfamily and the immunoglobulin-like domain of p140 using the ALIGN program (Protein Identification Resource, NBRF) indicate that the highest similarity scores (standard deviation > 3) are usually obtained with human or mouse T-cell receptor γ chains. The alignment of the immunoglobulin-like domain of p140 with the V region of human T-cell receptor γ -chain K20 (filename A24574) scored highest (s.d. = 5.41), with 25 per cent sequence identity and 10 per cent conservative substitutions. Similarity with s.d. > 4 is also observed with other superfamily members (such as human IgG γ 1 constant region GHU and rat Ig κ chain V region B23986). The carboxyl-terminal region of the immunoglobulin-like domain of p140 displayed the most marked sequence similarity with known immunoglobulin domains, including those in T-cell receptor α , β and γ V-chains (see figure).

The occurrence of several immunoglobulin-like domains in EBV-encoded proteins indicates a novel mechanism by which molecular mimicry might participate in multiple sclerosis (MS), which is

receptor antibodies would arise in certain predisposed individuals as a consequence of epitopes shared with viral proteins bearing immunoglobulin-like domains, such as EBV. Anti-T-cell receptor antibodies may be more effectively mitogenic in the central nervous system than in the periphery because of its low constitutive activity of complement and the lack of effective reticuloendothelial sequestration.

It is noteworthy that the murine homologue of a human T-cell receptor V β -gene

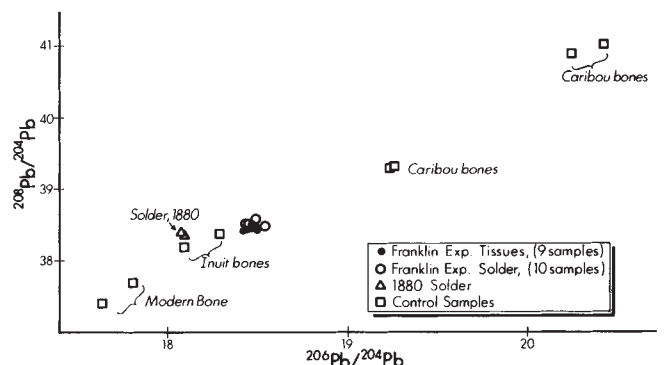
Did solder kill Franklins men?

SIR—One of the greatest mysteries of Arctic exploration is the loss of the 129 crewmen and officers of the Sir John Franklin expedition (1845–48) in search of a North-West Passage¹. As part of a forensic investigation of recently discovered human remains from the expedition, atomic absorption analysis yielded levels of lead in tissues consistent with acute lead intoxication². These findings indicate that it was exposure to toxic levels of lead that adversely affected the health, judgement and ultimate survival of the expedition members. It has been hypothesized^{2,3} that food preserved in soldered tins was the

source of the high lead levels. Because lead isotopes do not fractionate in biological systems, their ratios within human tissues are a reflection of the ratios from the contaminating source⁴. Matching lead isotope ratios would indicate that the lead in the human remains came from the soldered tins. We have carried out isotope studies of lead from the human remains and tins to test the hypothesis.

The materials used in the analyses were collected from Beechey and King William Islands, Northwest Territories, Canada. Bone samples from Inuit and caribou from the same time period and geographical

Lead isotope analyses, carried out by thermal ionization mass measurement using a VG-Micromass-30 mass spectrometer. The measured lead isotope ratios were corrected for mass fractionation, as determined by repeat analyses of US National Bureau of Standards NBS SRM 981 standard lead. The precision of the lead isotope ratios was established by replicate analyses and indicated a variation of one part per thousand at the 95% confidence level. Error limits are indicated (approximately) by the size of the circular symbols.



area, and with the same depositional characteristics, were analysed to evaluate the concentration and isotope composition of lead from the local environment. Samples of modern human bone and solder from food tins manufactured in the 1880s were included as additional controls.

The results of the lead isotope analyses are presented in the figure. The variation in the isotope ratios of the Franklin expedition samples is only slightly greater than the analytical uncertainty. The mean $^{206}\text{Pb}/^{204}\text{Pb}$, $^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{204}\text{Pb}$ ratios, with 1 σ uncertainty (for the tissue samples ($n=9$) were 18.46 ± 0.03 , 15.64 ± 0.01 , 38.48 ± 0.04 and for the solder samples ($n=10$) 18.46 ± 0.02 , 15.64 ± 0.01 , 38.50 ± 0.04). The slight spread in lead isotope values was expected as not all of the lead present in the tissues would have been acquired during the expedition.

Differential environmental exposures before the expedition would have assured a wide variety of isotopic ratios in the tissues of the various crewmen. The close clustering of isotope ratios, however, indicates that most of the lead in the tissues came from a single dominant source as the variability seen in the ratios is consistent with that observed when a small amount of lead from one source is mixed with a large amount of lead from another source; only a slight shift occurs in the isotopic ratios from that of the dominant contributor.

The isotope composition of lead in the tissues from the expedition members does indeed closely match that of the solder from the food tins. We conclude, therefore, that the toxic lead levels thought to be responsible for the demise of the expedition members were derived from the tinned foods.

WALTER KOWAL

OWEN B. BEATTIE

HALFDAN BAADSGAARD*

Departments of Anthropology
and Geology*,
University of Alberta,
Edmonton, Alberta T6G 2E3
Canada

PETER M. KRAHN

Riyadh Al Kharj Hospital
Programme,
PO Box 41726, Riyadh,
Saudi Arabia 11531

1. Beattie, O. *The Muskox* **33**, 68–77 (1983).
2. Kowal, W., Krahn, P. & Beattie, O. *Int. J. envir. analyt. Chem.* **35**, 119–126 (1989).
3. Beattie, O. In *The Franklin Era in Canadian Arctic History: 1845–59* pap. No. 131 (ed. Sutherland, P.) 141–148 (National Museum of Man, Ottawa, 1985).
4. Russel, R.D. & Farquhar, R.M. *Lead Isotopes in Geology* (Interscience, New York, 1960).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. They need not arise out of anything published in *Nature*. In any case, priority will be given to letters of less than 500 words and five references. □

The great fireworks illusion

SIR—The optical illusion (*Nature* **341**, 492; 1989) during pyrotechnic displays by which all exploding fragments appear erroneously to move toward the observer has a close parallel in the apparent motion of rotating searchlight beams.

In some cities, such as ours, powerful trailer-mounted searchlights are used to call attention to special sales and giveaways of merchandise when new retail stores open. Inclined searchlight beams stab the night sky as the light sources and reflectors rotate on a stable base. To an observer several miles away, the beams reaching into the sky appear to oscillate back and forth rather than to rotate.

The length of the lighted path in the sky appears longest when the vertical plane passing through the inclined beam is orientated at right angles to the view of the observer. As the inclined beam swings away from or towards the observer, the apparent length of the lighted sky path is foreshortened. The eye and brain consistently tend to interpret the observed foreshortening as a motion of the searchlight beam toward the observation point, rather than away from it.

As a result, the observer is prone to conclude that the motion of the searchlight is oscillatory rather than rotatory. The true motion is clear only if there is a low cloud cover. In that case, the searchlight illuminates a moving spot on the base of the clouds and the circular motion of the spot reveals the true rotatory motion of the beam.

I do not understand why this particular optical illusion should be so strong, but it is typically compelling without clouds to provide a reliable frame of reference.

WILLIAM R. DICKINSON

Department of Geosciences,
University of Arizona,
Tucson, Arizona 85721,
USA

SIR—Regarding the letter from J. D. Daniels (*Nature* **341**, 492; 1989), I have observed a similar illusion when driving at night: the rear light of the car in front sometimes appears to be approaching even when I do not subsequently catch up with that car. As with Daniels' firework example, this does not occur in daylight, but here there is little, if any, intensification of the light to give the illusion that it is 'getting larger', or drawing nearer. It also does not occur in heavy traffic, suggesting that the illusion is a function of the absence of local visual cues to depth. That other drivers who do not share my severe short sight seldom confirm the illusion would support this interpretation: in poor illumination they see faint cues better than I.

I suspect that this is a 'default option' of the human visual system. In the absence of

any good visual cues, objects are assumed to be approaching the observer. This would make good evolutionary sense: better assume that an object is approaching, and take appropriate action, than assume that it is stationary (or withdrawing) and take inappropriate inaction.

WILLIAM BAINS

Cambridge Laboratory,
PA Consulting Group,
Melbourn, Royston,
Hertfordshire SG8 6DP,
UK

SIR—The phenomenon pointed out by Daniels (*Nature* **341**, 492; 1989) becomes less surprising when its definition is appropriately changed. According to Daniels' definition, which I shall call analytical, there are several moving objects, each of which traces a predictable trajectory, but, surprisingly, some of the expected trajectories do not appear.

On the other hand, if we consider the firework as a single object, or as a growing sphere composed of self-luminescent fragments, it is possible to say that the observer does not distinguish the side of the sphere opposite to his cone of observation. I shall call this definition structuralist.

Levi-Strauss is said to have appreciated the value of structuralism by observing a dandelion fruit and considering that its form depends upon the arrangement of the pappi. There is a remarkable analogy between the forms of dandelions and fireworks. Clearly, partially empty spheres of this kind do not necessarily show the side opposite to the observer. A little experimentation will for example, show that whether the pappi of the opposite side of a dandelion are distinguishable depends on the distance of the observer from the flower.

Although with the analytical definition there is a 'magical' phenomenon to explain, the structuralist definition is an example of a common situation that needs clarifying. Thus the structuralist perspective turns the question about the illusion from strong to weak. This could be an example of how, in some circumstances, it is rational to maintain a certain level of complexity in an abstract reduction (structuralism, holism, system theory or, in this case, as perception is involved, also Gestalt psychology), despite the fact that there seem to be simpler elementary units. It is noteworthy that here a holistic perspective results in a more economic hypothesis than a reductionist one.

FRANCESCO PANSERA

Via Tevere,
20 Rome,
00198 Italy

Beyond the evidence

Duncan Campbell

The Myth of Heterosexual AIDS: How a Tragedy Has Been Distorted by the Media and Partisan Politics. By Michael Fumento. *Basic Books*: 1990. Pp. 411. \$22.95.

MICHAEL Fumento's writings and ideas have already made a significant public impact through the pages of the British and US newspapers in which he has argued his controversial case for more than two years. His analysis has been used to criticize general public-health education, to attack the scale and priority of AIDS research funding, and to undermine 'safer sex' education. The publication of his book unhappily coincides with renewed (and ultimately lethal) campaigns intended to assure orthodox heterosexuals that it is safe to have unprotected sex again, and to suggest to governments that they can stop spending money on AIDS because (for the most part) only gays face death.

Fumento is a polemicist, not a scientist. His fundamental hypothesis, only slightly disguised, is that AIDS is necessarily a disease of disgrace, requiring the transmission of both stigma and a virus in order for someone to become infected. Fumento's analysis of HIV transmission is derived from this proposition. Hence, in Fumento's world, heterosexual transmission of HIV cannot and does not occur if the infected partner was himself or herself infected with HIV through heterosexual intercourse. The fact that such transmission is both theoretically predictable from coefficients of transmission measured

in partner studies, and is observed epidemiologically — on a particularly large scale in Africa — causes him little discomfort. Where he cannot argue such evidence away by anecdote or by a declamatory sideswipe at "alarmists", he passes on, leaving the rents in the fabric of his argument unpatched.

The true myth of this book — the proposition that heterosexual transmission of the 'pure' sort (which he labels tertiary transmission), cannot occur — is an invention by Fumento. Inevitably, he ties himself in knots. First, he admits that his title is misleading: "The 'myth' of heterosexual AIDS consists of a series of myths, one of which is *not* that heterosexuals do get it." Fumento then re-asserts his central axiom, and claims that heterosexuals get HIV infection "only from shared needles, from transfusions, from clotting factor, . . . from their mothers at or before birth, and sometimes through sexual intercourse with persons in these categories and with bisexuals" — but never from another heterosexual, infected heterosexually. The axiom is restated again and again: "AIDS will pick off a person here and there in this group [heterosexuals], but the original infected partner *will be* in one of the two groups in which the disease is epidemic" (my emphasis).

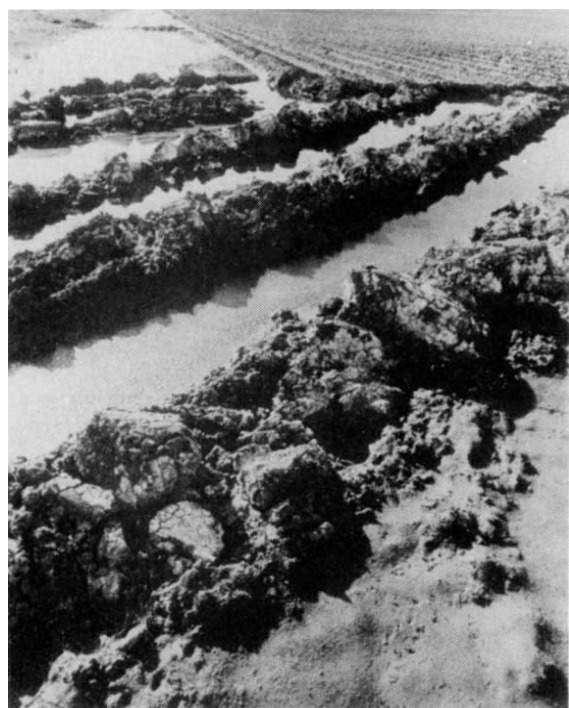
This proposition is untenable and

untrue, and the author knows it. Heterosexual men can infect heterosexual women with HIV, he admits, and heterosexual women can infect heterosexual men. Transmission of HIV can occur in a single (penetrative) sexual act. The consequence is of course that HIV can pass sexually from man to woman to man to woman, all heterosexuals. Fumento's axiom is the real myth. He admits this: "There is no physical law saying that tertiary transmission cannot occur."

Fumento goes on to admit that tertiary transmission has been observed, then makes several extraordinary attempts to argue away data showing heterosexual transmission which he finds inconvenient. There is a well-known report, published in 1986, of a Swedish sailor, Lars, found to be the centre of a network of six entirely 'tertiary' cases of heterosexual HIV transmission. "Possible explanations for Lars", writes Fumento, include "an extremely efficient strain of HIV, his proclivity for sex in a manner unusually conducive to HIV transmission" (this is code for anal intercourse) or "a lesson in the laws of probability". By his last option, the author means to suggest that the heterosexual infection of the sailor, three of his subsequent female partners, the husband of one of these women and the child of another was merely a coincidental statistical artefact. According to Fumento's axiom, the sailor — being heterosexual himself — cannot transmit HIV to another heterosexual, therefore these other infections are coincidences, not transmission pathways. He rejects the clear and obvious deduction from these and similar data.

It may be thought that anyone whose attitude to empirical evidence is as cavalier as this has little to contribute to public debate except the aggravation of prejudice and ignorance. The one important, substantiated part of Fumento's 400-page tract could have been said in a long footnote elsewhere. This is where Fumento decries some of the more hysterical propagandists who have portrayed AIDS as a new Black Death leading to widespread and general morbidity. The main offenders in this category include such authors as Masters, Johnson and Kolodny, whose text *Crisis; Heterosexual Behaviour in the Age of AIDS* was published in 1988, and propagandists like Dr. John Seale, who continue to urge the public to fear the possibility of social and/or casual transmission of HIV.

But Fumento's criticism of these writers degenerates into lumping together all those whose views he does not care for, labelling them abusively as "alarmists" or "condomites". In fact, Fumento has much in common with authors like Seale, including a desire to enlarge his reputation for AIDS iconoclasm, and a conservative political agenda in which homophobia plays a major part. ►



Losing ground — reproduced from *Cadillac Desert* by M. Reisner, the photograph shows salt deposits covering farmland in the San Joaquin Valley, the sparse natural resources have been stretched, exploited and depleted and a million acres of the American west in California alone may be affected. Published by Secker and Warburg in Britain, £25; Viking/Penguin in the United States, hbk \$22.95, pbk \$9.95.

Such views often also infect official HIV statistics, and have done so for some time. British government AIDS and HIV statistics are a case in point. The published statistics contain single cumulative totals of homosexual and drug-use HIV infections, but offer no cumulative total for heterosexual transmission. Instead, sub-totals are given for heterosexual partners of bi/homosexual men or drug users, for others "possibly infected abroad", for "no evidence of exposure abroad" and "undetermined". One does not have to master deconstructionism to see the slant in this breakdown, or that it follows the thinking behind the Fumento myth. The message is that the ordinary heterosexual who avoids partners in Africa or in so-called 'risk groups' isn't likely to get infected by HIV.

This message is no accident. The sub-categories in these statistics were devised in order to minimize the impact of the discovery of the first (UK) heterosexual cases of HIV transmission. One senior government epidemiologist at the time openly spoke of the disease being, in his view, restricted to "bum-bashers"; ergo, like Fumento, he attempted to argue from such a proposition that heterosexual cases of HIV infection were being reported in error and should be recategorized.

Fumento is a little less despising of homosexuals and less orientated towards conspiracy theories than other writers of his genre. But his personal agenda emerges clearly at the beginning and at the end of his book. Only a writer whose prejudices deny humanity could write in such bad taste as this: "Although AIDS is no joke, there is good news and bad news about the length of HIV infectiousness . . . the 'good news' [is] that the great majority and perhaps almost all, of HIV-infected persons will develop debilitating symptoms or die". Puns about sodomites ("condomites") and the prurient repetition of accounts of exotic gay sexual activities demonstrate that Fumento has his own sexual axe to grind.

His summary chapter is entitled *The AIDS Lobby — are we giving it too much money?* His answer, clearly, is yes: "it is not fair to penalise victims of cancer and other life-threatening illnesses because they do not knit quilts or blockade the Golden Gate Bridge". There is, Fumento claims, "a misallocation of funds and fear resulting from the 'democratisation' of AIDS" away from identification with gay men. Fumento's essential idea is thus that you can't catch HIV unless you are exposed to stigma, as he sees it, as well as the virus. It is an unscientific, mischief-making hypothesis. It is wrong, and thus deadly. □

Duncan Campbell is Associate Editor, New Statesman/Society, Foundation House, Perseverance Works, 38 Kingsland Road, London E2 8DQ, UK.

Welfare state

Michael F.W. Festing

Animal Research and Ethical Conflict. By M.T. Philips and J.A. Sechzer. Springer-Verlag: 1989. Pp.251. DM132, £43.50.

Animal Experimentation: The Consensus Changes. Edited by Gill Langley. Macmillan: 1989. Pp.260. Hbk £27.50; pbk £10.95. Distributed by Chapman and Hall in the United States \$45, \$15.95.

THE use of animals in research has engendered strong emotions for at least two hundred years. Increasingly restrictive legislation, and an ever more sophisticated animal-welfare lobby, are trends that can no longer be ignored by scientists. *Animal Research and Ethical Conflict* is an analysis of the literature (books, journal articles and news items) on laboratory animal welfare published in the United States between 1966 and 1985. More than half of the book is reference material, including a useful summary of international legislation on laboratory animals.

The historical introduction provides some interesting nuggets of information. I knew that Adolf Hitler was an antivivisectionist, but had not realized that in 1933 "vivisection of animals of whatever species" was prohibited in Prussia. The analysis of more than 700 articles reveals some interesting trends, interspersed with some acerbic comments by the authors. Thus "scientists cannot ignore legislation or other regulations that restrict their work, but they can and did for a long time ignore the ethical questions raised by critics", and "morality remains a subject that most scientists are wary of . . .".

In the United States, the implementation of animal-welfare legislation depends critically on institutional animal care and use committees (IACUCs), rather than on professional inspectors as is the case in the United Kingdom. IACUCs will have to balance animal protection with scientific progress, and the authors conclude that a substantial effort in training in the humane use of animals, ethics, and potential alternatives is needed. This book can be recommended to all who are concerned with the use of laboratory animals.

Animal Experimentation: The Consensus Changes is a collection of essays by eleven authors who are professionally concerned with the welfare of research animals. Topics covered include ethics, evidence for pain and suffering in animals, critiques of current methods, alternatives, and methods of improving animal experiments when they cannot be avoided.

Some of the arguments are controversial. A chapter by Tom Reagan, a philosopher, concludes that we have no moral right to harm or even kill animals. The implications of this need further discus-

sion, given that we would soon be overrun by rats and mice if we took this conclusion too literally. The use of animals in toxicology is criticized by several authors, Erik Millstone from the University of Sussex states that in his opinion "... well-conducted relevant animal experiments . . . provide the lesser of a set of evils . . .", but he concludes that "while animal [toxicity] tests enable us to learn how commercial chemicals affect mice, rats, rabbits and guinea-pigs, we know very little about their probable effects on humans". In Britain, at least, he also considers that the criteria by which the results of animal studies are judged are often inconsistent and commercially biased.

Efforts to develop alternatives are discussed by Martin Stephens, who draws attention to Russell and Birch's "high-fidelity fallacy". A high-fidelity model is one which closely resembles humans, but low-fidelity models may have better discriminating power, and hence be more valuable. Many *in vitro* tests are likely to have high discriminatory power, and their use should continue to be explored in more detail.

Occasionally, the bias of one of the authors shows up. Robert Sharpe claims that animal experiments have made no contribution to medical progress. Thus "... Louis Pasteur, whose apparently successful attempts to develop a vaccine against rabies . . . had glamorised . . . animal experiments . . ." is written off by claiming that "clinical experience has shown that few people bitten by a rabid animal actually develop the disease so those who miraculously 'recovered' after a course of inoculations may not have been infected in the first place". The fact is that we do have anti-rabies vaccines and these have been developed as a result of Pasteur's work. By attempting to claim otherwise, Sharpe undermines his own credibility.

If there is to be a dialogue between research scientists and those concerned with animal welfare (and I consider this to be very worthwhile), there should be no suspicion that either party is being economical with the truth. However, scientists who claim all is well do not have a monopoly on the truth. Gill Langley in her excellent essay "Plea for a Sensitive Science" gives some specific examples where "... a reluctance on the part of some scientists to be open minded and apply fair standards to new methods of research and testing" has resulted in the unjustifiable use of whole animals. This is reinforced by Clive Hollands' chapter "Trivial and Questionable Research on Animals". In conclusion, this book should be compulsory reading for all those engaged in animal research. □

Michael F.W. Festing is at the MRC Toxicology Unit, Carshalton, Surrey SM5 4EF, UK.

Pooled resources

Robin Dunbar

The Evolution of Social Systems. By John Paul Scott. *Gordon and Breach: 1989. Pp.355. Hbk \$90; pbk \$49.*

EXPLAINING the evolution of social systems continues to remain one of the chief challenges in biology. The issues involved are immensely complex, so that despite significant advances in recent years, we are still a long way from solving all the problems posed by the diversity of social systems in the natural world.

In his book, Scott, who has had a long and distinguished career as a behavioural geneticist, attempts to tackle this formidable problem. Much of Scott's concern is with what he sees as the genetic naivety endemic to this whole area. The burden of his argument is that the simple one-gene/one-character models that so often lie at the heart of evolutionary analysis are oversimplifications to the point of absurdity: most genes have more than one effect and most characters are determined by more than one gene. What is missing in our conventional approach to these problems, he suggests, is a systemic view.

Scott's contention is that the organism is not only a tightly integrated genetic and physiological system, but also that it operates as a component in a series of macroscopic systems of ever-greater inclusiveness (such as the social group or the ecosystem). The systemic nature of evolutionary processes becomes particularly intrusive with the evolution of two sexes: the propagation of genes is now no longer simply a matter of frenetic genic replication, but rather one of compatibility between the two parents and their respective sets of genes. The conventional wisdom that evolution occurs as a result of competition is, Scott insists, a direct result of our Western competition-orientated capitalist society.

Scott's thesis seems to rest on several misunderstandings that will surprise, perhaps even alarm, evolutionary biologists. He insists, for example, that the fundamental concept of fitness refers to the ability to survive. Although it undoubtedly did so in Darwin's lifetime, evolutionary biologists now conventionally define fitness in terms of relative rates of gene replication, survival of the individual being but one of its many components. This view, however, leads Scott to compound his problems: he insists that, because entities other than the gene can survive more or less successfully, so selection must occur at levels other than the gene (including at the very least the cell, organism, group and ecosystem). Unfortunately, this is to conflate a crucial distinction between the level at which

selection falls and the entity whose propagation is thereby selected for. The former can indeed involve many different layers in the biological cake but, with very rare exceptions, the latter is always the gene.

Scott goes out of his way to paint a fair and balanced picture of the theories he considers. Unfortunately, his discussion of sociobiology is marred by two serious misunderstandings. First, he assumes that the term competition refers to direct fighting between individual organisms. In fact, competition in an evolutionary context always refers to competition between genes in a gene pool. Contrary to his apparent supposition, competition at the genic level commonly gives rise to co-operation between individuals. Indeed, mutualism is widely recognized as the primary factor underlying the social system of most higher vertebrates.

The other misunderstanding is Scott's assumption that sociobiology is a theory about the ontogeny of behaviour. This is a surprisingly common mistake and it is time that it was finally laid to rest. Sociobiological explanations are concerned with function, and even though function in an evolutionary context is necessarily concerned with the replication of genes, this in no way implies that behaviour is genetically determined. To suppose that

it does is to confuse questions about function and ontogeny.

Although I have considerable sympathy with Scott's attempt to emphasize the systemic basis of many biological processes, he has, I fear, missed a golden opportunity to demonstrate the value of a systems approach. Instead, he has opted to argue in favour of a complete genetic description of behaviour. This is tantamount to rewriting the whole of chemistry in terms of quantum physics. Although a perfectly feasible project, a computer capable of handling the calculations has yet to be built, and in the end would probably reveal little more than is already known about the processes of chemistry. Even if this were a promising research programme in evolutionary biology, our knowledge of genetics is still too rudimentary; no discipline should commit itself so irrevocably to a particular version of another's. Doing so lays the parasitic discipline's entire research programme open to premature demise when new developments in the host discipline lead to major episodes of theory replacement. □

Robin Dunbar is in the Department of Anthropology, University College London, Gower Street, London WC1E 6BT, UK.

Well tuned

R. H. Kay

Hearing: Physiological Acoustics, Neural Coding, and Psychoacoustics. By W. Lawrence Gulick, George A. Gescheider and Robert D. Frisina. *Oxford University Press: 1989. Pp.409. £35, \$39.95.*

INSTANT reprints of the proceedings of auditory conferences abound. That contrasting species, a mature general text on

hearing, is rare. Some textbooks are flawed by excessive research-group parochialism or by the hasty inclusion of work too new to have been properly assessed. *Hearing*, avoiding such traps, gives comprehensive coverage from peripheral biophysics to sensation. It is scholarly, with an unashamed tutorial emphasis that should please its intended readership of maturing undergraduates and aspiring neuroscientists.

Modern information is included but placed in context. Notable discoveries during the past decade on the micro-

Scanning electron micrograph of nerve fibres crossing the organ of Corti in the human ear.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Science Photo Library

structure of hair cells, histochemistry and intracellular biophysics are described with reference to the active processes undoubtedly needed for cochlear tuning. There is a welcome emphasis on mammalian work. Most students of audition are interested in the human condition and deserve not to be unduly distracted by the intriguing comparative neurophysiology of turtles, toads and lizards. Neural mechanisms are so divergent among vertebrate classes that direct extrapolation between them is hazardous.

'Sound-localizing' systems of the central nervous system are fully discussed. This term commonly means direction-finding alone, and ignores distance, the extra parameter needed for true localization. Here, the authors discuss the ambiguity in binaural disparities arising from distance and direction. They also clearly state that one-eared human direction-finding is better than can be explained by neurophysiology. Happily, they avoid using the bat as a model: that auditory specialist exploits a personal sound generator to compute distance by timing, a method denied to other terrestrial mammalian species.

The authors also touch on the discrepancies between conclusions from

neurophysiological experiments on putatively 'free-field' (as in the open air) direction-finding. Laboratory experiments usually proceed in far from strictly anechoic conditions, thus are not truly free-field, and very small animals are used, so that measuring/monitoring equipment near their heads introduces significant diffraction, which is hard to calibrate. In 1938, the careful work of Stevens and Newman on human direction-finding necessitated sitting the subject on top of a pole on a Harvard rooftop in the quiet of the night to satisfy free-field conditions. Unsurprisingly, neurophysiologists today do nothing like that. Yet for this work the only effective substitute is probably a true anechoic room. These expensive constructions are very rare, used often for commercial or governmental testing but normally unavailable for academic research. This is probably why discrepancies arise.

Hearing handsomely serves its intended readership. Physicists, engineers, musicians and philosophers should also find it enjoyable and informative. □

R. H. Kay is Emeritus Fellow in Physiology of Keble College, Oxford OX1 3PG, UK.

Chronicle

Charles Tanford

Companion to the History of Modern Science. Edited by R. C. Olby, G. N. Cantor, J. R. R. Christie and M. J. S. Hodge. Routledge: 1989. Pp.1081. £65, \$85.95.

FORMAL professional education in science suffers from a deplorable neglect of the historical element, and many a working scientist has a genuine desire to remedy this deficiency. The scientist's own life is dominated by the sequence of experiments in his laboratory (or the sequence of theoretical explorations in his head) — a succession of sudden insights, false leads, unexpected results and so on. Did his predecessors in his own field report similar struggles in their memoirs or letters? Can he learn from their experience? From a broader perspective, when and how were the now solid foundations of his field (or of science in general) established? When, for example, was the need for conservation laws in physics first recognized? Who first discovered the true function of the liver and how did it happen? Can the scientist see his own work as contributing to some such historic growth of knowledge?

Where should the scientist turn for answers to such questions? To the historian of science, one might naturally suppose, but regrettably it's not that simple. For 'history of science' has become an academic discipline in its own right,

posting its own exclusive claims in the territories of knowledge and scholarship — with emphasis on originality, with specialized jargon and with in-house controversies. Its dominant themes are about science and its history, not for their own sake, but as fuel for complex historiography. Is science just another example of a "customarily expressed belief", its vaunted appeal to experimental test no more than a myth? Do popular theories for scientific revolutions ("internalist" versus "externalist") provide valid models? Mere factual history is banal or so, at least, this *Companion* would have us believe. "There is much more to the history of science than a chronicling of who got what right when", the *Companion* tells us and, a little later, "it [the history] is quite clearly much too important to be left to the self-interested and distorted perceptions of the working scientists themselves".

The *Companion* defines 'modern' science as beginning in the sixteenth century and the editors have by choice excluded medicine and technological applications. The book contains 67 essays, written by 61 contributors. Fewer than half the essays actually mention 'facts' of history at all — who did what or the traditional history of changing ideas — and, of course, most of those that do so continue with comment on that history in theoretical terms. Robert Olby, for example, writing about the molecular revolution in biology, moves quickly from fact to theory: "It is tempting to depict modern molecular biology as an example of a

Kuhnian phase of revolutionary science leading to the establishment of a new paradigm. Twentieth-century science can then be neatly packaged into two major revolutionary phases — first quantum physics, then molecular biology." Is this a trivialization of what is arguably the most important event in the entire history of biology or is it a new and deeper view of it?

The overall coverage of the *Companion* is, intentionally, selective rather than comprehensive. Almost every essay ends with valuable notes and with a long list of references for further reading. Contributions are organized into several different categories. One group adds up to a kind of course of instruction in related disciplines (philosophy, sociology, marxism, linguistics), encouraging historians to extend the scope of their work in those directions. Contributions that focus on particular subdivisions of science are divided into "turning points" (ranging from Copernicus to cybernetics) and "topics and interpretations" (dealing with longer-term matters). A final group of about twenty essays ("themes") goes far afield to issues that are outside the mainstreams of both traditional and analytical history — science and war, science and imperialism, science and education, scientific professionalism and so on. I found this to be the least satisfactory part of the book, with many glaring omissions of what surely should have been cardinal topics. One essay in this group, by R. M. Young, is pure political venom — sweatshops in the Korean computer-chip industry seen as a direct consequence of the scientific worldview and its adoption by capitalist society.

The *Companion* obviously speaks chiefly to the professional historian, for whom it will surely be a welcome reference, the first broad compendium of its kind. As in any multi-authored book, however, there is a lack of uniformity. Some of the essays are relatively non-technical and provide good reading for the general reader. For example, Robert Fox on the rise and fall of laplacian physics between 1799 and 1815; Raymond Fancher on Sigmund Freud, an objective account with references to a wide variety of interpretations; John Stachel on relativity; J. J. Gray on geometry and space; and Ernan McMullin on philosophy of science between 1600 and 1900.

Even the more technical parts might profitably be perused by the much denigrated laboratory scientists, most of whom are probably ignorant about current trends in 'history of science' as formally defined. Given that historians reflect how posterity may see today's science in the future, it is valuable to be made aware of the distinct ideology that is emerging in the field. □

Charles Tanford is at Tarlswold, Back Lane, Easingwold, North Yorkshire YO6 3BG, UK.

Holography in artificial neural networks

Demetri Psaltis, David Brady, Xiang-Guang Gu & Steven Lin

The dense interconnections that characterize neural networks are most readily implemented using optical signal processing. Optoelectronic 'neurons' fabricated from semiconducting materials can be connected by holographic images recorded in photorefractive crystals. Processes such as learning can be demonstrated using holographic optical neural networks.

IN recent years there has been a marked resurgence of interest in artificial neural networks¹. The structure and the principles of operation of such 'neural' computers are to a large extent biologically motivated. The reason for attempting to derive clues from neurobiology to build computers is the sharp distinction that exists between the types of problems for which modern digital computers are useful and the tasks at which humans and other animals excel. The classic example is the distinction between pattern recognition and arithmetic: humans do arithmetic poorly, but easily outperform computers in the simplest recognition tasks. Research in neural computing is based on the premise that this difference in capabilities is due to basic differences in the hardware and the ways in which the hardware is programmed. Although in some instances, particularly in the case of sensory systems that are relatively well understood, it may be possible to construct circuits that are reasonably accurate replicas of biological systems, more often one extracts only the basic properties evident in the nervous system and uses this information to guide the design of the computer. The hope is that if one is successful in identifying truly relevant properties, then the neural analogy should provide a valuable contribution to the design process even though the circuitry of the nervous system is not understood in detail.

The rapid progress in modern computer design is owed in part to the separation of algorithmic and architectural issues from the physics of the devices used to implement the algorithm; in this way, a programmer may develop algorithms without detailed knowledge of the physics of the semiconductor hardware. Unfortunately, this separation is difficult to maintain in highly parallel architectures, such as neural networks. One of the main issues in parallel computing is the relationship between the properties of a parallel architecture and the efficiency with which it solves problems. The physics of the hardware is a central consideration in this debate. The efficiency of communication between nodes of the network is a major limiting factor in hardware design. In parallel electronic circuitry most of the system area must be devoted to interconnecting wires². This imposes limitations on the hardware architecture and therefore solutions have been sought that go beyond the traditional silicon technology. We will discuss one such solution, holographic optical interconnections³, in the context of the most highly interconnected parallel architecture, artificial neural networks⁴⁻¹¹.

Basic principles

The principles that are most commonly used in the design of neural computers are as follows.

Massive parallelism. If we think of an individual neuron as a computational element, then the human brain consists of 10^{11} – 10^{12} such elements. For the execution of a task that involves the visual pathway, a large portion of these units are active. This is in sharp contrast to a conventional digital computer, where perhaps a few thousand out of more than 10^8 transistors are active at any one time during a typical computation.

Dense interconnections. A neuron in the brain typically receives inputs from several hundred to several thousand other units to produce its own output, which is broadcast to roughly the same number of units. It is widely believed that the synaptic strengths

serve as part of the memory of the system. This connectionist view has been adopted almost universally in the design of artificial neural networks. Connectionist networks achieve high computation speed through specialization by encoding general knowledge about the computational problem in the structure of the network.

Learning. The ability of the nervous system to perform desirable computations accurately is acquired through the genetic code and adaptation. The ability of a conventional digital computer to perform useful tasks derives from the design of its circuit architecture (the hardware) and its programming (the software). We can think of the hardware design as being analogous to the role of genetics and the software to adaptation. There is, however, a sharp distinction between programming and adaptation. A programmer explicitly designs every step of the computation as a sequence of precise mathematical statements. The neural approach to computation develops the required software for tackling a particular problem through a sequence of trials during which the parameters of the system are adapted to achieve a desired goal. The success of such learning procedures is critically dependent on the learning algorithm and the hardware architecture. The system designer must make both of these choices *a priori*. Even though these choices are typically very difficult to make, they are relatively few. In practical terms, for certain problems the learning approach can provide a method for transferring some of the burden of programming the computer from the user to the computer.

The architecture of optical neural networks

Figure 1a is a schematic diagram showing the various components of a pair of neurons and the way in which they are

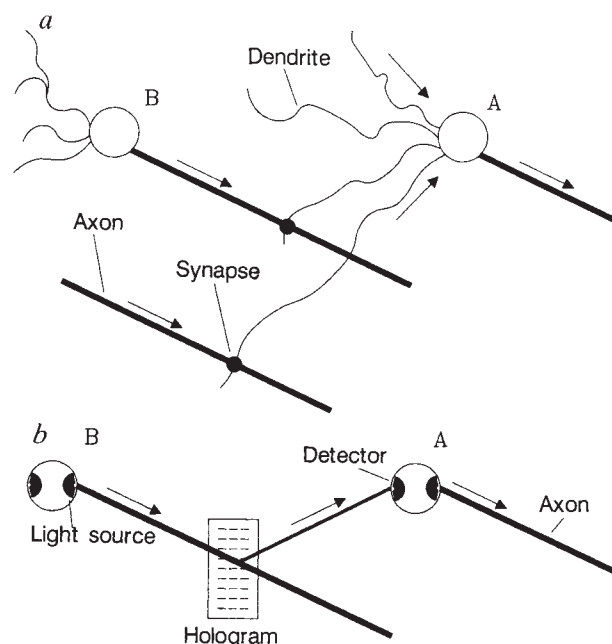


FIG. 1 Conceptual structure of *a*, a neural system and *b*, its optical analogue. See text for details.

connected. Each neuron receives inputs through the synapses on its dendrites, it processes these inputs in some fashion, and then it broadcasts the result on the axon where it is picked up by the dendrites of other neurons. This description is somewhat simplified (for example, dendrites can be two-way channels¹²), but it is a view that has been adopted almost universally in neural-network models. The diagram in Fig. 1b shows the holographic analogue of the two neurons in Fig. 1a. The output of each neuron is a light beam, and the activity of the neuron is coded in the amplitude or intensity of the optical signal. The input of each neuron is a light detector which senses the amount of light that is directed towards it. A holographic grating is placed in the path of the output beam of neuron B, which diffracts the incident light. The direction of the diffracted beam is determined by the period and the orientation of the grating. With an appropriate holographic grating, light from neuron B illuminates the detector of neuron A. In this way a signal is generated in A as a result of the activity in B, and we say that the hologram connects the two neurons. The strength of the connection can be modified by adjusting the modulation strength of the holographic grating. There are some direct analogies between the component structure of a neuron and its optical simulation. The output light beam has the role of the axon, broadcasting the signal from each neuron. The holographic grating plays the part of the synapse, directing the signal from one neuron to the next, and the optical pathways along which light is transferred from the hologram to the detector area of the neuron are analogous to the dendrites. The device consisting of the optical beam generator, the detector and the circuits that process the detected signal is reminiscent of the soma of the neuron. In real neurons some processing tasks can take place on the branches of the dendritic tree, but in the optical simulation all integration and computation tasks are concentrated on this integrated unit, which we refer to from now on as the 'neuron'.

Having established the basic structure of the optical analogue of a single neuron, the next issue is the way in which we may connect many such units together to form optical neural networks. Suppose that the third neuron is to be connected to neuron A as well. This can be optically simulated by superimposing a second holographic grating in the same crystal that was used to store the interconnection pattern between the first two neurons. This second grating has a distinct spatial orientation and period and it is tuned to redirect light from the third neuron onto the same input (the detector) of neuron A. This is a second difference between the real neuron and its holographic realization: the synapses are not localized at the intersection between axon and dendrite, but are implemented instead in a distributed manner, each sharing the entire volume of the holographic medium.

Distributed connections have advantages and disadvantages—the disadvantage compared to the localized implementation is the reduced control of individual synapses. The adjustment of the strength of one synapse may inadvertently affect other synapses as well. Accommodating this limitation is perhaps the most crucial research issue in this field and we discuss this further below. The advantages of the distributed holographic synapses are high storage density and ease of fabrication. The number of distinct synapses that can be packed in a hologram of volume V is V/λ^3 , where λ is the wavelength of the light¹³. This corresponds to 10^{12} synapses per cm^3 for the typical operating wavelength $\lambda \approx 1 \mu\text{m}$. There are other factors, primarily relating to the physical properties of the material used, that can prevent us from realizing this upper limit. In practice it is possible to realize 10^9 – 10^{10} interconnections per cm^3 . The distributed holographic storage of the weights is also the ultimate in simplicity in terms of device fabrication: it involves only crystal growth.

Although holograms can simulate the function of the synapses, the axons and the structure of the dendritic tree, optoelec-

tronics are used to perform the nonlinear computations of the neuron and to provide the optical energy needed to incorporate such devices into large networks. Typically, the neuron has a 'soft threshold', which is to say that the relationship between the summed post-synaptic signals into a particular unit and the output of this unit is an S-shaped curve. To implement this nonlinearity we must provide a mechanism by which the incoming signals can interact. Optical signals do not interact in free space, so a material is introduced at the location of the neuron that senses the presence of the light and, in response, changes its electronic state. This modifies the optical properties of the material and an optical beam travelling through this medium can therefore be affected by the presence of another beam. Various potentially appropriate devices based on optical nonlinearities are under development^{14–16}, but at present the most effective method for achieving such nonlinear optical effects involves detection of the optical signals, followed by nonlinear processing of the electronic signals and subsequent regeneration of optical signals through a light modulator or source. Semiconductor materials are particularly well suited for this purpose because it is possible to fabricate monolithic devices that incorporate all three functions. The unit shown in Fig. 2 is an example of a 10×10 array of optoelectronic neurons fabricated in gallium arsenide¹⁷. Each unit consists of a pair of transistors, a photodetector and a light-emitting diode (LED). An incident signal above threshold is sensed by the photodetector amplified by the transistors and rebroadcast by the LED. It is possible to construct devices with $\sim 10^4$ neurons per cm^2 that can be switched in parallel at rates of over 10,000 Hz. Such a capability, although not readily available at present, can be obtained by rather straightforward advances.

The common feature of the various types of optical and optoelectronic neuron arrays is planar layout. This is a consequence of the techniques used in fabrication, namely epitaxial growth or deposition of a sequence of thin layers on a two-dimensional substrate. As a result, we obtain two-dimensional arrays of neurons or neural planes. This technological constraint leads to an optical neural network that consists of planes of neurons separated by a space in which reflection holograms specify the connection between units in a single layer and transmission holograms interconnect neurons in two different layers. Although this structure of layers of neurons separated by interconnections emerges from technological constraints of the optical implementation, it is nevertheless similar to the structures used in almost all neural-network models.

The three-dimensional structure of the optical architecture has a potential advantage over electronic implementations. In electronics, the neurons (simulated by transistors) and the interconnections (implemented with wires, resistors and transistors) are both fabricated on the same planar surface¹⁸. Typically, the area devoted to the interconnections is the largest fraction of the available area and this severely limits the size of the network that can fit on a single chip. In the optical implementation, the active planes can be populated with neurons alone because the interconnections are realized holographically using the third dimension. As a result, the number of neurons that can fit on a plane of a given area is at least C times larger for the optical implementation, where C is the average number of connections per neuron. To make the comparison another way, the electronic implementation requires at least C times more chips to form a network with a given number of neurons. C can be well over 100, so optical connections provide a large savings in the number of semiconductor devices needed. The relative disadvantage of optics is that, with only a few exceptions, the technology is at an early stage. Two basic components are needed: the neural planes and the holographic interconnections. We have discussed briefly how optoelectronics can provide the answer for fabricating the neurons and hereafter we focus on the holographic interconnections.

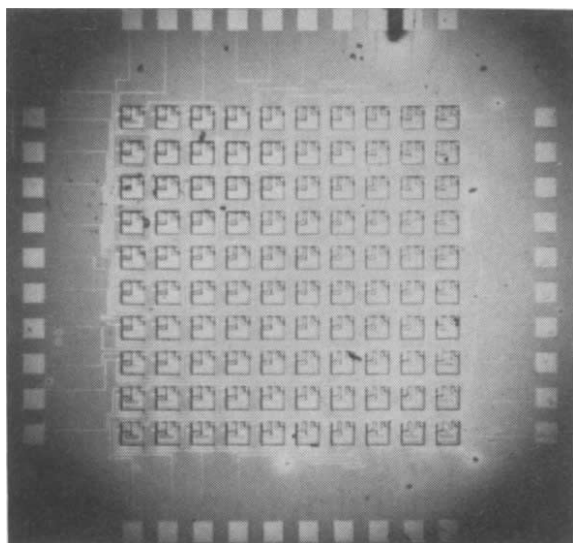


FIG. 2 Photograph of an optoelectronic implementation of a 10×10 array of neurons in gallium arsenide. The area of the array is 5×5 mm.

Holographic interconnections

Figure 3 shows a particular realization of a holographic neural-network architecture¹⁰. Each resolution element (or pixel) at the first neural plane can be a location for a neuron. The light from a pixel is collimated and diffracted by a holographic grating. The diffracted light is focused by a lens onto a pixel at the output neural plane. We now discuss how, given the location of two neurons at the input and output planes, the appropriate grating can be formed. As shown in Fig. 3, the portion of the light from the input neuron that is not diffracted by the hologram is reflected by a phase-conjugate mirror (PCM). A PCM, unlike a conventional mirror, reflects a collimated beam back along its original path. A PCM is a nonlinear crystal that forms a hologram of the incident beam. The recorded hologram is then illuminated from the opposite direction, resulting in a reconstructed beam that precisely retraces the path of the beam incident on the PCM. Light from the output neuron propagates towards the left in Fig. 3 and is collimated by the lens. This beam and the beam that is reflected by the PCM form a sinusoidal interference pattern at the hologram. The interference pattern exposes the hologram and a sinusoidal grating is recorded. This arrangement ensures that the grating will interconnect the input and output pixels or neurons that were used to record it. The strength of the connection is controlled by varying the exposure during recording. The connection is stronger when the light is brighter or the exposure time longer. If x_i and x_j are the amplitudes of the light from the input and output neurons, respectively, then the change in the strength or the weight of the connection formed by the grating is $\Delta\omega_{ij} \propto x_i x_j$. x_i and x_j can be, in general, bipolar signals. When the sign of x_i matches the sign of x_j then ω_{ij} increases; otherwise it decreases. Alternatively, the system can be designed with x_i and x_j being positive only. In this case $\Delta\omega_{ij}$ can be only positive, and separate excitation and inhibition channels are constructed. Because the algorithm causes only positive changes in the

weights, they will eventually exceed the operating range of the holographic medium and the medium will be saturated. This can be avoided by introducing a weight decay mechanism. Both modes of operation are possible, but here we consider only the case of bipolar neurons.

We now consider connections between multiple neurons at the input and output planes. If N is the number of pixels in one dimension for both input and output planes, there are N^2 pixels in each plane. If we use each of these pixels as a site for a neuron, and we want to interconnect independently each of the input neurons to all the output neurons, we need N^4 interconnections, which implies that N^4 weights or gratings must be stored in the hologram. This poses the question of the maximum number of gratings that can be distinguished in a hologram of volume V . Assuming that the resolution of the holographic medium in any direction is $\delta > \lambda$, the total number of resolution cells or distinct samples in the volume of the holographic crystal is V/δ^3 . From the sampling theorem, the number of distinct sinusoids, or gratings, that can be recorded in this medium is equal to the number of samples, so the maximum number of independent interconnections that can be supported by a hologram of volume V is V/δ^3 . This guides the design of specific architectures. As an example, we consider a network in which neurons in two adjacent layers are arbitrarily interconnected. The product of the number of units in the first plane, $N_1 \leq N^2$, and the number of units in the second plane, $N_2 \leq N^2$, cannot exceed the available holographic interconnections in the crystal:

$$N_1 N_2 \leq \frac{V}{\delta^3} \quad (1)$$

If this relationship is violated then an attempt to establish the interconnection between one pair of input and output neurons automatically specifies the interconnections for other pairs as well. To avoid this we can either increase the volume of the crystal or decrease the density with which the neural planes are

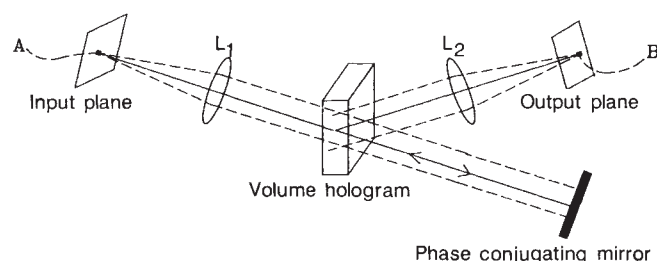


FIG. 3 Holographic neural-network architecture. The light amplitude at each pixel on the input plane corresponds to the output of a neuron. The signal detected at each pixel on the output plane corresponds to the summed inputs to a neuron. The phase-conjugating mirror reflects the input exactly back on itself. Light propagating back towards the input interferes with control signals generated at the output plane to adaptively update the interconnection pattern stored in the volume hologram.

populated (that is, decrease N_1 and N_2). Here we will use the second strategy. In particular, we assume that the hologram is a cube, with edges of length L , and $L/\delta = N$. In other words, the number of resolvable points in any one dimension is the same for both the neural planes and the hologram. This is a special case, but it is also the most sensible way to design such optical systems. In this canonical system, the total number of gratings available in the hologram is $V/\delta^3 = (L/\delta)^3 = N^3$. If the density of neurons remains constant from layer to layer ($N_1 = N_2$), then from equation (1) the maximum number of neurons per plane is $N^{3/2}$. The remaining task is the selection of the correct $N^{3/2}$ pixels from the N^2 available sites at each plane for the placement of neurons.

Figure 4 shows three examples of patterns that appropriately sample $N^{3/2}$ points out of N^2 for $N = 16$ (refs 19, 20). In each case the larger circles represent neuron locations on the regular two-dimensional grid of dots. The number of neurons in each case is $16^{3/2} = 64$. We derive these patterns using a process of elimination. Each time we attempt to add a new neuron to the input (output) sampling grid, we check to see whether this new neuron is already connected to one of the neurons selected previously at the output (input) by an existing grating. If it is,

we eliminate the position of the new neuron from the sampling grid; if it is not, we select this position for a new neuron, which implies that gratings are established to connect this new neuron to all the neurons now selected at the output (input). In addition, we eliminate all the positions at the output (input) to which this new neuron is connected by existing gratings. We arrive at the complete sampling grids by systematically iterating this procedure. The patterns that result are not unique. Different patterns emerge by adopting different methods for selecting the next available location. The most systematically derived and regular patterns are shown in Fig. 4a. We created the pattern in Fig. 4c by borrowing a concept from fractals. These sampling patterns can be thought of as fractals with dimension $\frac{3}{2}$. If we replace each neuron in any of the patterns in Fig. 4 with the entire pattern itself, we end up with a larger pattern of the same fractal dimension. We created the patterns in Fig. 4c by starting with the patterns of Fig. 4b, drawn for $N = 4$, and then following the above procedure. Interestingly, new patterns created in this manner have the property that each holographic grating interconnecting an input and output pair on these samplings grids is distinct if the seed pattern from which it was generated satisfies this same property.

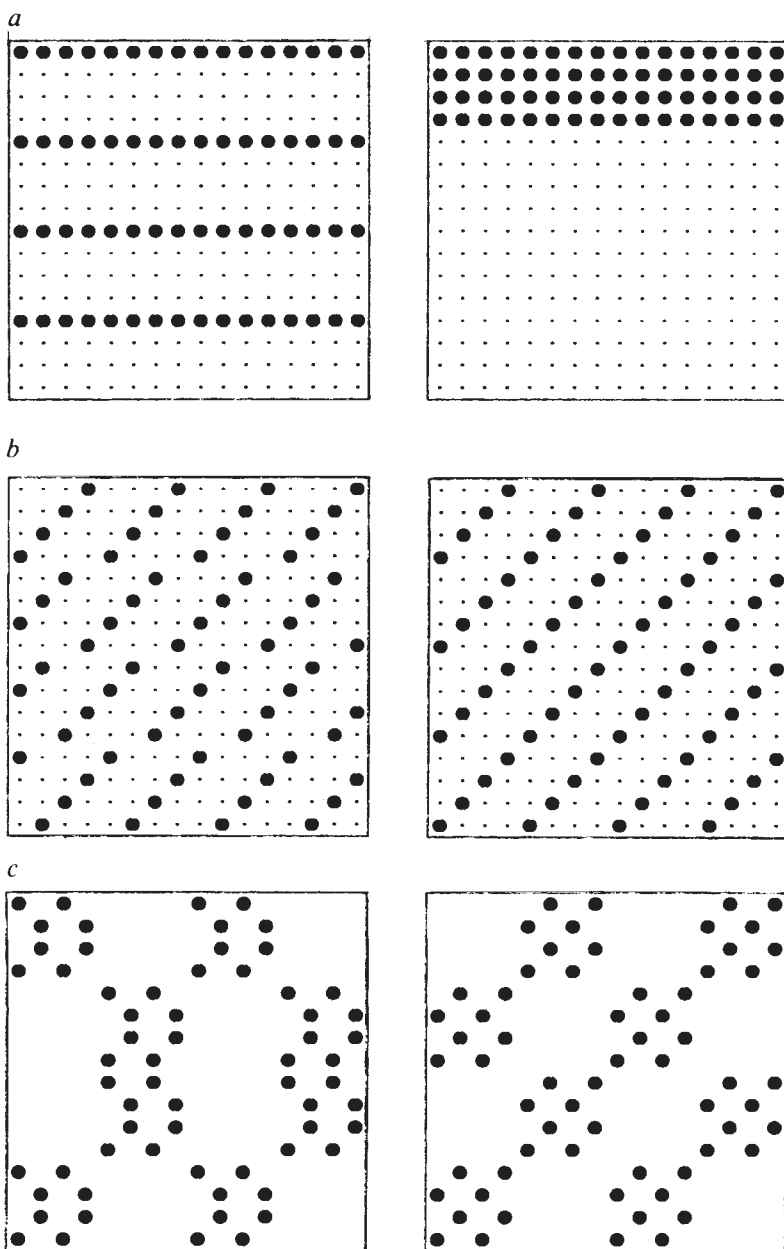


FIG. 4 Nondegenerate sampling grids. Each pair of sampling grids is such that the holographic grating that interconnects any point on the input grid with any point on the output grid is unique to that input-output pair.

The necessity of using the fractal sampling grids becomes apparent when we use the architecture of Fig. 3 as an associative memory. The goal is to record the appropriate holographic gratings in the crystal such that for a particular pair of images, the light at the output neural plane is image B if the illuminating pattern at the input is image A. We attempt to record the hologram by setting the state of the input neural plane to be the image A and the state of the output, image B (ref. 21). The input neural plane emits light towards the right in Fig. 3 and the output emits light towards the left. The light from the input is back-reflected by the PCM and interferes with the light from the output to form the hologram. The recorded hologram is reconstructed by placing pattern A at the input and the reconstruction forms at the output. The photographs in Fig. 5a show the result that was obtained when this procedure was performed in the laboratory. We recorded the holograms in a photorefractive lithium niobate crystal using an argon-ion laser. The two patterns used in the experiment were the capital letters A and B. The reconstruction obtained when the A was placed at the input is a smeared B. We can think of the holographic associative memory described above as an interconnection of each point on the letter A to all the points on the letter B. If this intercon-

nection pattern is in fact recorded in the hologram, then when the A is present at the input, all the neurons that make up the letter A will produce the same reconstruction of B at the output. The result is a strong reconstruction of the letter B. In general, the output will not respond unless the activation at the input plane is sufficiently similar to the pattern used for recording the hologram. In the experiment the letters A and B were drawn on a regular rectangular grid. As a result, the establishment of interconnections of input and output points on the letters A and B inadvertently forms connections between points on the letter A at the input and points on the letter B at the output, thus giving the smeared reconstruction of B in Fig. 5a. When we repeated the same experiment but sampled the patterns on the fractal grids of Fig. 4b, we obtained the sampled patterns shown at the top of Fig. 5b. The reconstruction sampled by the fractal grid at the output, shown at the bottom of Fig. 5b, is now perfect.

This experiment vividly demonstrates how the sampling grids permit us to use each of the gratings recorded in a volume hologram to establish a distinct interconnection between two neurons. The high density with which gratings can be stored makes it possible to construct very large, densely connected networks. The capability to build large networks is useful, however, only if it is accompanied by the capability to fully load information in the connections of the network. In this experiment only a single association was stored in the hologram. In theory, this holographic associative memory has the storage capacity to support a number of associations between pairs of images equal to the number of neurons at each plane. We discuss the procedures for achieving this capacity next.

Learning in photorefractive holograms

The recording mechanism described in the previous section for establishing a connection between a single pair of neurons is very similar to hebbian learning, in which the strength of the connection between two neurons is reinforced if their activation patterns are correlated. The holographic associative memory is an extension of this basic learning mechanism, with the connection between sets of input and output neurons reinforced simultaneously and in parallel. Almost all of the learning algorithms that have been developed for neural-network models make use of this basic mechanism at the synapse level. The algorithms differ primarily in the procedures used to establish the activation of the neurons during training, but the modification of each weight during a training cycle is almost universally a simple product of activations of the two neurons that are connected by the weight. Therefore, at least in principle, holographic interconnections can be used for all such learning algorithms. Most learning algorithms require a very large number of training cycles. Here we describe how photorefractive crystals can be used to form holograms recorded by an arbitrarily long sequence of exposures, thereby extending the applicability of holographic interconnections to a broad class of learning algorithms.

Photorefractive crystals are a class of nonlinear optical materials that combine three properties: they are photoconductive (light causes current to flow), electro-optic (the presence of an electric field modifies the index of refraction), and they have defects (or 'traps') in their lattice that can be optically ionized²². Holograms are recorded in these crystals by exposing them to light with a photon energy that is matched to the energy required to ionize a trap. The recorded hologram is stored in the spatial distribution of the ionized traps in the crystal. When we make a sequence of exposures, the distribution is continuously rearranged but the material imposes no limitation on the number of exposures. This is in contrast to photographic film, for instance, where each exposure causes an irreversible chemical change and ultimately the material saturates. Although there is no limit in the number of exposures for a photorefractive crystal, we must still design a sequence of exposures that can load the appropriate weight values in the finite pool of trap sites that are available.

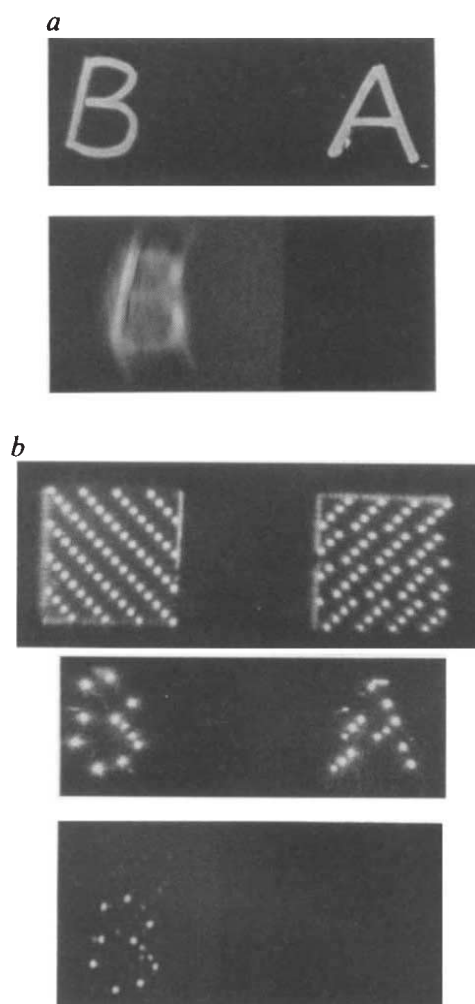


FIG. 5 *a*, Photographs of the letters B and A carried on an Ar⁺ laser beam and the reconstruction of the hologram made by the interference of these two patterns. The hologram is reconstructed with A and yields a smeared version of B. The smearing is due to degeneracies in the hologram. *b*, Photographs of a pair of nondegenerate sampling grids, the letters B and A as sampled by these grids, and a reconstruction of a hologram made with the sampled patterns. The hologram is reconstructed with the letter A and yields a high-fidelity image of the sampled letter B.

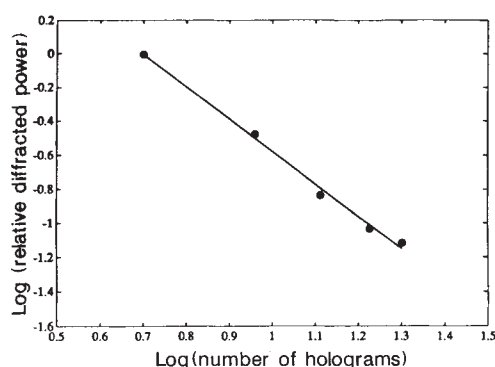


FIG. 6 Relative diffraction efficiency against the number of holograms stored for the fifth association stored between a name and a code. The slope of the least-squares fitted line is -1.95 .

We will describe the extension of the holographic associative memory to multiple associations, as a simple example. We use again the basic arrangement of Fig. 3 with a photorefractive crystal used as the holographic recording medium. To establish the holographic connections that will associate M input and output image pairs, we superimpose a sequence of holograms. Each association is recorded in precisely the same way as before but the length of each exposure is different for each association. The first association is exposed for a long enough time to use all the available traps in the crystal for recording this one hologram. When the crystal is exposed to the second association, the resulting redistribution of the charge among the traps causes the first hologram to be partially erased so the second is recorded. We choose the length of the second exposure so that the two recorded holograms have equal strength. The subsequent exposure records the third association until all three have equal amplitude, and so on. At the end of M exposures, the hologram contains M associations, all with equal strength. This exposure schedule results, however, in an average diffraction efficiency for each association proportional to M^{-2} (ref. 11) rather than the ideal M^{-1} which would be obtained if the M holograms fully occupied the available dynamic range.

We can demonstrate the storage of multiple associations in a single hologram using a simple experiment. The patterns we associated were the first names of people affiliated with our research group and random dot patterns. Using the set-up in Fig. 3, the sampling grids in Fig. 4a and the exposure schedule above, 20 exposures were made in a strontium barium niobate photorefractive crystal. Once the holograms were recorded, each of the random dot patterns could be used to retrieve its associated name (or vice versa). We used a liquid-crystal spatial light modulator²³ to simulate the neural planes. The logarithm of the diffraction efficiency of each hologram is plotted as a function of the logarithm of the number of exposures in Fig. 6. The predicted M^{-2} relationship agrees well with the experimental results.

The reduction in efficiency that accompanies the increase in the number of exposures, ultimately limits the number of associations that can be superimposed on a single hologram. The limit is reached when the strength of the reconstruction of an individual association becomes comparable to the noise in the

system. Mok *et al.* recently recorded 500 holograms of visual scenes on a single crystal using this technique²⁴. This is probably close to the practical limit of the method. The number of training cycles required for the implementation of error-driven algorithms, such as error back-propagation, is not known *a priori* and it can easily be more than a thousand. As each training cycle corresponds to a holographic exposure, the present method must be extended so that a sequence of exposures of arbitrary length can be done. We are now experimenting with a method that involves two photorefractive crystals, serving as the short-term and long-term memory of the system²⁵. Holographic exposures are accumulated in the short-term memory and its contents are periodically copied to the long-term memory. The long-term memory is also periodically rejuvenated by copying its contents into the short-term memory and back again. Our initial experiments confirm that this continuous exchange of information between the short-term-memory and long-term-memory crystals results in a non-decaying hologram for arbitrarily long training sequences.

Conclusion

There are remarkable analogies between the basic properties of neural-network models and simple holography. These similarities make it possible to construct, in the laboratory, very large densely connected multilayer networks with relative ease. At present, two obstacles prevent the practical use of such systems. The first obstacle is the lack of practical devices for the simulation of the neural planes; several candidate devices are, however, being investigated. The second obstacle may be more serious: we do not yet have sufficient knowledge, either from observation of the nervous system or from theoretical models, of the operation of large complex networks. Although recent progress has engendered a great deal of optimism for short-term practical benefits, we are, most probably, only at the beginning of a long journey. □

D. Psaltis and S. Lin are at the Department of Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA. D. Brady is now at the Beckman Institute, University of Illinois, Urbane, Illinois 61801, USA. X.-G. Gu is now at Rockwell International, 1045 Camino dos Rios, Thousand Oaks, California, 91360, USA.

- Anderson, I. A. & Rosenfeld, E. (eds) *Neurocomputing* (MIT Press, 1988).
- Sutherland, I. E. & Mead, C. A. *Scient. Am.* **237** (Sept.), 210-228 (1977).
- Goodman, J. W., Leonberger, F. J., Kung, S. Y. & Athale, R. A. *Proc. IEEE* **72**, 850-866 (1984).
- Psaltis, D. & Farhat, N. H. *Opt. Lett.* **10**, 98-100 (1985).
- Abu-Mostafa, Y. S. & Psaltis, D. *Scient. Am.* **256** (March), 88-95 (1987).
- Farhat, N. H. *Appl. Opt.* **26**, 5093-5103 (1987).
- Owechko, Y., Dunning, G. J., Marom, E. & Soffer, B. H. *Appl. Opt.* **26**, 1900-1910 (1987).
- Anderson, D. Z. *Opt. Lett.* **11**, 56-58 (1986).
- Yariv, A. & Kwong, S. K. *Opt. Lett.* **11**, 186-188 (1986).
- Wagner, K. & Psaltis, D. *Appl. Opt.* **26**, 5061-5076 (1987).
- Psaltis, D., Brady, D. & Wagner, K. *Appl. Opt.* **27**, 1752-1759 (1988).
- Kuffler, S. W., Nicholls, J. G. & Martin, A. R. *From Neuron to Brain* (Sinauer, Sunderland, 1984).
- Van Heerden, P. J. *Appl. Opt.* **2**, 393-400 (1963).
- Gibbs, H. M. *Optical Bistability: Controlling Light with Light* (Academic, New York, 1985).
- Smith, S. D., Walker, A. C., Tooley, F. A. P. & Wherrett, B. S. *Nature* **325**, 27-31 (1987).
- Miller, D. A. B. in *Electro-optic and Photorefractive materials* (ed. Günter, P.) 35-49 (Springer, Berlin, 1987).
- Kim, J. H., Lin, S. H., Katz, J. & Psaltis, D. *Proc. Soc. Photoinstrum. Engners* **1043**, 44-52 (1989).
- Mead, C. A. *Analog VLSI and Neural Systems* (Addison-Wesley, Reading, Mass., 1989).
- Psaltis, D., Yu, J., Gu, X. G. & Lee, H. *2nd Topical Meet. Opt. Computing* (Opt. Soc. Am., 1987).
- Lee, H., Gu, X. G. & Psaltis, D. *J. appl. Phys.* **65**, 2191-2194 (1989).
- Gabor, D. *IBM J. Res. Develop.* **13**, 156-166 (1989).
- Günter, P. & Huignard, J.-P. (eds) *Photorefractive Materials and Their Applications I and II* (Springer, Berlin, 1988).
- Bleha, W. P. *et al. Opt. Engng* **17**, 371-384 (1978).
- Mok, F. *Opt. Soc. Am. A. Mtg., 1989 Technical Digest Series* **12** (1989).
- Brady, D., Hsu, K. & Psaltis, D. *Opt. Lett.* (submitted).

ACKNOWLEDGEMENTS. The authors thank H. Lee, K. Wagner, K. Hsu, J. Kim, S. Li, J. Yu and Y. A. Mostafa for their contributions and comments. S.L. is supported by an SDIO/ISTC Fellowship. D.B. acknowledges the support of the Aerojet Corporation. This research is supported by DARPA, the Air Force Office of Scientific Research, and the Army Research Office.

Mechanical basis for low-angle normal faulting in the Basin and Range province

H. J. Melosh

Lunar and Planetary Laboratory and Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

Low-angle normal faults seem to violate the criterion that relates fault dip to stress orientation. To explain such faults one must postulate a 45° rotation of principal stress directions at mid-crustal depths, which is difficult to understand on the basis of previous elastic crust models. Such a stress rotation is shown to be a natural consequence of a low-viscosity lower crust.

DETAILED geological mapping in the Basin and Range of the western United States has recently revealed a type of fault that, at first sight, defies a simple mechanical explanation. Low-angle normal faults or detachments of large areal extent have been recognized in many areas of the Basin and Range province¹⁻³. The large extent and extremely shallow dip of these low-angle normal faults, or detachment surfaces, indicates that the faults originally formed at low angles⁴. Where exposed, the footwall is generally foliated to mylonitic⁵⁻⁷, indicating ductile strain at mid-crustal depths. Although the hanging wall presently exhibits a more brittle deformational style, low-angle slip probably began at mid-crustal depths near the brittle-ductile transition.

The problem with low-angle normal faults is that observations on rock fracture indicate that most faults form as shear fractures at orientations of $\pm 30^\circ$ to the maximum compressional stress axis. Anderson⁸ argued that, near the Earth's surface, one principal stress direction must be perpendicular to the surface because shear stresses vanish at a free surface. Depending on whether this perpendicular principal stress is the maximum, minimum or intermediate stress, the resulting faults are either normal, thrust or strike-slip with dips of 60°, 30° or 90° (vertical) respectively. The only way low-angle normal faults can form as shear fractures is if the principal stress directions rotate rapidly between the surface and depths ranging from 6 to 10 km where low-angle normal faults are believed to form. The alternative, that low-angle normal faults represent a new type of fracture in which the slip plane is nearly perpendicular to the maximum compressive principal stress, is not supported by any experimental evidence on rock fracture of which I am aware. But it seems equally difficult to understand how, in a broad terrane of relatively uniform crustal thickness, principal stress directions can rotate by 30° or 45° between the surface and mid-crustal depths. Where do the large shear stresses necessary to cause the rotation come from? Recent models that treat the crust as an elastic plate^{9,10} have shown that large principal-axis rotations are possible in the vicinity of large compressive, flexural or shear loads where shear stresses change rapidly, but such explanations imply special circumstances that involve hidden sub-crustal forces or structures.

A more general explanation of stress rotation in the Basin and Range can be derived from the rheological structure of the crust in this high heat-flow region. Figure 1a shows the effective viscosity η_{eff} as a function of depth in a crust and upper mantle extending at a mean strain rate of $2 \times 10^{-15} \text{ s}^{-1}$ (assuming 100% Basin and Range extension over 15 Myr; C. Chase, personal communication) in a model crust similar to that of ref. 11. A 10-km-thick upper crust is rheologically identical to westerly granite, a 10-km-thick lower crust acts like Maryland diabase, and the upper mantle is dominated by olivine. Although some

aspects of this model may be unrealistic (the effect of varying water content is not addressed here, and the transition from upper crust to lower crust is probably gradual, thus reducing the viscosity peak at the top of the lower crust), the general pattern is believed to be accurate. The mechanical implication of this structure is shown in Fig. 1b, where the Maxwell time, $\tau = \eta_{\text{eff}}/\mu$, is plotted as a function of depth (μ is shear modulus). The crust responds elastically to loads applied for time intervals shorter than the Maxwell time, but behaves as a viscous fluid for loads applied for longer periods. In the Basin and Range, only the upper few kilometres of the crust behave elastically for loads that exist for more than ~ 1 Myr, and in the lower part of the crust the Maxwell time is only $\sim 10,000$ yr. These short times are due to both the relatively high heat flow of the Basin and Range, as rock viscosity is strongly temperature-dependent, and also to the relatively high extensional strain rate, as the effective viscosity also decreases with increasing strain rate (see, for example, ref. 12).

In the Basin and Range, the crust and upper mantle can be thought of as a thin upper layer, which behaves elastically, grading into a lower viscous layer, which has one or more additional elastic layers underneath. Such a crust can clearly

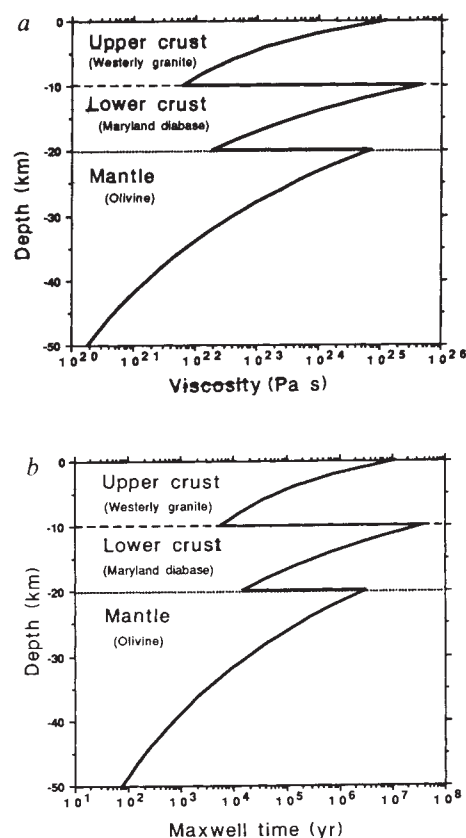


FIG. 1 Rheological structure inferred for the present Basin and Range using the thermal gradient of Smith and Bruhn¹¹ and the rheological data of Kirby and Kronenberg^{12,17} for the materials indicated; a, shows the viscosity as a function of depth and b, shows the Maxwell time. Elastic moduli in the crust are from Dziewonski *et al.*²².

not be treated as a uniform, homogeneous elastic slab. This structure is the key to understanding the principal-stress rotations apparent from low-angle normal faults. Here I show that if the upper elastic layer is displaced horizontally with respect to a lower elastic layer, the intervening viscous layer is subjected to simple shear. Such displacement is a natural result of localized faulting in the upper crust, opposed to uniform stretching in the lower elastic layer(s). The principal stress directions in the viscous layer thus become orientated at 45° to the vertical, consistent with the formation of low-angle ductile faults—by the Mohr-Coulomb failure law, brittle rocks with angles of internal friction near 30° fail along planes at angles of $\pm 30^\circ$ to the compressive stress axis, but in ductile failure the angle is expected to be $\pm 45^\circ$. Experimental confirmation of this expectation is, however, sketchy (T. and J. Tullis, personal communication). The principal-stress rotation between the surface and mid-crustal depths is thus due to the change in rheology from elastic to viscous at these depths, coupled with some differential slip between the layers.

Analytical model of extension

It is easy to show that the principal stress directions in a viscoelastic layer sheared between two elastic layers approach an orientation of 45° to the vertical within a few Maxwell times of the application of the shear. The initial stress state of any part of the crust is generally complicated because it depends on the history as well as the present geological circumstances of that particular piece of crust. However, in general, within the constraints of keeping the analysis simple, suppose that in some region the vertical stress is $\sigma_{zz} = \rho g z$, where z is depth, g the acceleration due to gravity, and, for simplicity, we assume that the density ρ of the crust is uniform. Let the initial horizontal stress be $\sigma_{xx} = k\sigma_{zz}$, where k is a constant in accordance with observation¹³, and the shear strain rate $\dot{\epsilon}_{xz} = \dot{\epsilon}_0$ is constant. Solution of the Maxwell viscoelastic constitutive relations in conjunction with the stress equilibrium equations shows that the time evolution of the stresses is

$$\sigma_{xx}(t) = \sigma_{zz} + (k-1)\sigma_{zz}e^{-t/\tau} \quad (1a)$$

$$\sigma_{zz}(t) = \rho g z \quad (1b)$$

$$\sigma_{xz}(t) = \sigma_{xz}(0)e^{-t/\tau} + 2\eta\dot{\epsilon}_0(1 - e^{-t/\tau}) \quad (1c)$$

where τ is the Maxwell time, $\tau = \eta/\mu$, η is viscosity and μ is the elastic shear modulus.

Within a few Maxwell times, these stresses approach a steady state in which $\sigma_{xx} = \sigma_{zz} = \rho g z$, $\sigma_{xz} = 2\eta\dot{\epsilon}_0$. The angle θ between the vertical and the maximum compressive principal stress is

(equation 2-34 in ref. 14)

$$\tan 2\theta = \frac{2\sigma_{xz}}{\sigma_{xx} - \sigma_{zz}} \quad (2)$$

as σ_{xx} approaches σ_{zz} the right-hand side of this equation approaches infinity and θ therefore approaches 45° , as described (see Fig. 2). A somewhat similar idea was proposed by Bradshaw and Zoback¹⁵, who investigated the refraction of principal stress directions between layers of differing viscosity, although they did not discuss viscoelastic stress relaxation or make it clear that differential slip across the layers is necessary. The above result corresponds to theirs in the limit of infinite viscosity contrast (that is, elastic layer against viscous layer).

Simple shear of a viscoelastic layer thus results in a pure shear stress state within a few Maxwell times. The memory of any initially different stress state is completely lost as viscous relaxation proceeds. The final state of stress at depth is thus consistent with low-angle ductile shear faults, whose failure planes are located at $\pm 45^\circ$ to the principal stress directions, according to the Mohr-Coulomb construction. The problem of explaining the occurrence of low-angle normal faults in the Basin and Range thus reduces to one of determining how the uppermost elastic layer can slip relative to one of the deeper-lying elastic layers, as the above analysis shows that when such slip occurs the principal stress axes are always rotated in the intervening viscoelastic layer.

Consider the simple model of Basin and Range extension shown in Fig. 3. An elastic crustal block of length L and thickness h_{eff} is underlain by a viscous lower crust of thickness h_{lc} . Beneath this lower crust lies stiffer upper mantle which is under sufficient pressure that it flows in a ductile (not viscous) fashion when stresses within it exceed its yield strength, extending at a uniform strain rate $\dot{\epsilon}$. The coordinate system is chosen such that the velocity of the upper mantle relative to the upper crust is $v_b = \dot{\epsilon}x$, so that the centre of the crustal block is stationary with respect to the upper mantle, but the differential velocity increases linearly with increasing distance from the block's centre. This is the situation expected for an extending mantle beneath a brittle crust that cannot support large strains without fracture. The fractures are represented by vertical breaks at $x = \pm L/2$ in this simple model. The stress acting across these breaks is set equal to a constant, σ_{xx}^{fr} , which is less than the breaking strength of the crust. The extensional strain rate in the brittle crustal block is essentially zero: the regional strain imposed by the extending upper mantle is accommodated by relative motion of crustal blocks at their edges, where the relative velocity is $\Delta v = \dot{\epsilon}L$.

The viscous lower crust in this model, $-h_{\text{lc}} < z < 0$ is thus

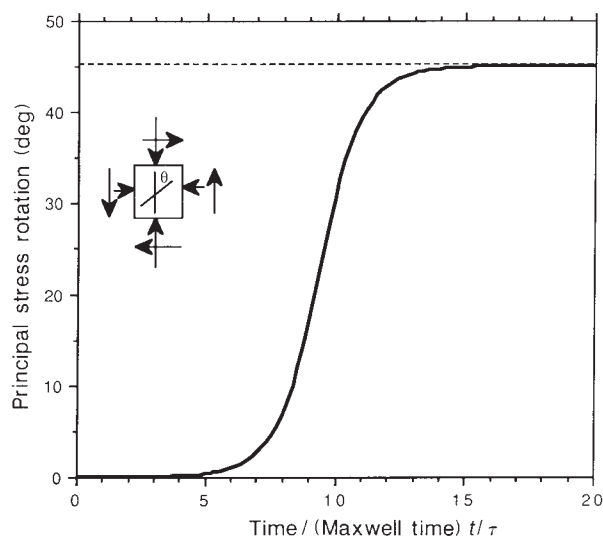
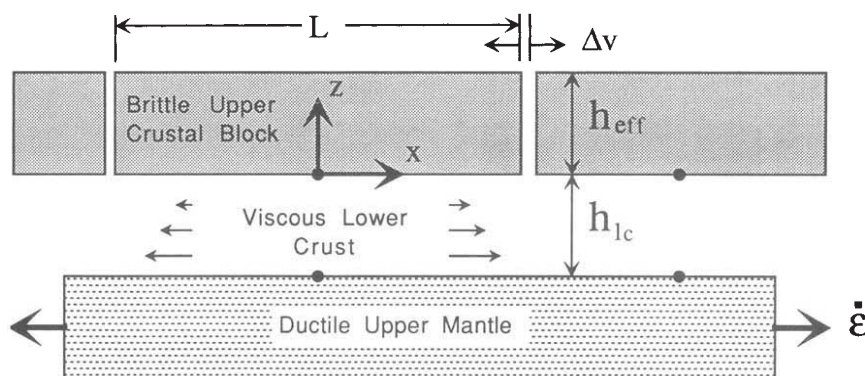


FIG. 2 Rotation of principal stress axes in a viscoelastic material as a function of time, derived from equations (1) and (2). After a period of ~ 10 Maxwell times the principal stresses rotate to an angle of 45° . The inset shows the definition of rotation angle θ and the short arrows illustrate the applied stresses.

FIG. 3 Schematic illustration of the analytical crustal-block model for Basin and Range extension. An effectively rigid upper-crustal block is separated from a ductile extending upper mantle by a viscous lower crust.



subject to a differential velocity between its top, which is in contact with the brittle non-extending crust and its bottom, which is in contact with the uniformly extending mantle. The components of the velocity field in the lower crust are determined by the boundary conditions and the assumption that the horizontal velocity gradient is linear (good for a thin lower crust):

$$v_x(x, z) = -\dot{\epsilon} \frac{xz}{h_{lc}} \quad (3a)$$

$$v_z(x, z) = \dot{\epsilon} \frac{z^2 - h_{lc}^2}{2h_{lc}} \quad (3b)$$

where v_z is determined from v_x and the condition of incompressibility, $\nabla \cdot \mathbf{v} = 0$. As $h_{lc} \ll L$, in general $v_z \ll v_x$: the z -component of velocity corresponds to a uniform thinning due to extension that is much slower than the horizontal stretching. The viscous (non-lithostatic) stresses developed in the lower crust may be calculated from the velocity field by the usual constitutive relations ($-h_{lc} < z < 0$),

$$\sigma_{xx}^{lc} = 2\eta \frac{\partial v_x}{\partial x} = -2\eta \dot{\epsilon} \frac{z}{h_{lc}} \quad (4a)$$

$$\sigma_{zz}^{lc} = 2\eta \frac{\partial v_z}{\partial z} = 2\eta \dot{\epsilon} \frac{z}{h_{lc}} \quad (4b)$$

$$\sigma_{xz}^{lc} = \eta \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) = -\eta \dot{\epsilon} \frac{x}{h_{lc}} \quad (4c)$$

Note that all the stresses are scaled by the same factor of viscosity multiplied by the strain rate, $\eta \dot{\epsilon}$, so that the principal-stress rotation angle, being a ratio (equation (2)), is independent of either viscosity or strain rate. At the top of the lower crust, $z = 0$, σ_{xz}^{lc} is the only non-zero stress, and the principal stress axes are rotated to 45° here. Furthermore, throughout most of the lower crust $z \ll x$, so the principal stress axes are rotated close to 45° everywhere except near the centre at $x = 0$.

The shear thus developed in the lower crust exerts a basal shear stress σ_b on the brittle upper crust, where $\sigma_b = 2\eta \dot{\epsilon}_{xz}$

$\eta \dot{\epsilon}(x/h_{lc})$. By the stress equilibrium equations, this basal shear results in an extensional stress in the overlying brittle lithosphere¹⁶ because the shear stress must decrease linearly to zero at the free surface at the top of the upper crust. The non-lithostatic stresses in the brittle upper crust are thus ($0 < z < h_{eff}$)

$$\sigma_{xx}^{uc} = \frac{\eta \dot{\epsilon}}{2h_{eff}h_{lc}} \left[\left(\frac{L}{2} \right)^2 - x^2 \right] + \sigma_{xx}^{fit} \quad (5a)$$

$$\sigma_{zz}^{uc} = 0 \quad (5b)$$

$$\sigma_{xz}^{uc} = \frac{\eta \dot{\epsilon}}{h_{lc}} \left(1 - \frac{z}{h_{eff}} \right) x \quad (5c)$$

where σ_{xx}^{fit} is the stress at the ends of the crustal block. The horizontal stress is smallest at the ends of the crustal block and maximum in the middle. If this stress were to cause the crustal block to break up further, the site of the next fracture would thus be in its middle. The shear stress in the crust declines from σ_b at the bottom to zero at the top, as it must. The vertical stress is zero (the lithostatic stress must be added to the above stresses to get the full stress field, but because the lithostatic stress plays no part in controlling principal stress direction it can be safely neglected here). Equations (5a-c) show that throughout most of the upper crust $\sigma_{xx}^{uc} \gg \sigma_{xz}^{uc}$, because $L \gg h_{eff}$, so that principal-axis rotations are small, and normal faults would be expected to form in the usual 60° dip orientation in this region.

The principal stress directions for this model are shown in Fig. 4. The length of the unfractured crustal block is 200 km and the thickness of the upper and lower crust are 6 km and 14 km, respectively. No other parameters, such as viscosity or strain rate, enter into the determination of the stress directions. It is clear that in the vicinity of the ends of the crustal blocks the orientations of the bounding faults should undergo a sharp change as they cross the upper/lower crust boundary, changing abruptly from the usual 60° dip in the upper crust to nearly flat in the lower crust (assuming that faults in the ductile zone form at $\pm 45^\circ$ to the compressional axis, as predicted by Mohr-Coulomb theory). Of course, the upper/lower crustal boundary is actually gradational, being established by the decrease in

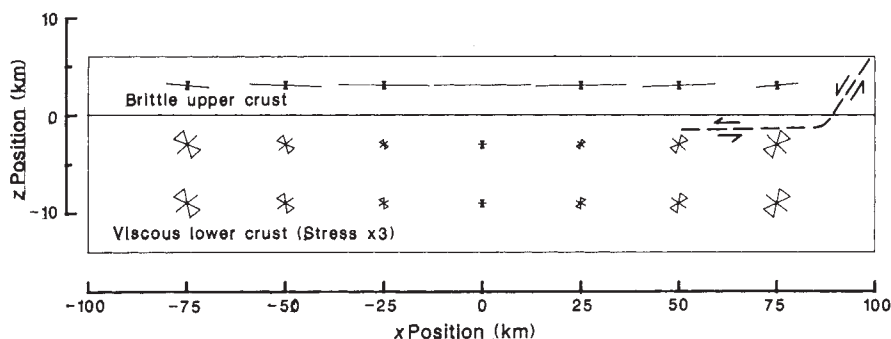


FIG. 4 Stresses in the analytical crustal-block model derived from equations (5a-b). The triangular symbols represent compressional principal stresses, the lines are extensional principal stresses, and the symbol size is proportional to the size of the stress. Note the exaggeration of the symbol size in the lower crust. The dashed line illustrates the geometry of a fault expected to form under the influence of this stress field.

Maxwell time with depth, so the orientation of the fault planes should rotate from steep to nearly flat over some depth interval, but the rate at which the Maxwell time declines with depth (Fig. 1b) is so high that this interval will be at most a kilometre or two.

Finite-element model of extension

Although the above analytical model displays most of the behaviour necessary to illustrate stress rotation in the lower crust, it may be argued that the vertical boundaries of the edges of the crustal blocks are unrealistic, that the measured rheology of crustal rocks is non-newtonian, and that the boundary between upper and lower crust is artificially chosen to lie at a fixed depth. To answer this criticism, I also present the results of a finite-element computation designed to be as realistic a representation of the Basin and Range crust and upper mantle as possible.

The basic structure of this plane-strain finite-element model of the Basin and Range extension is that of an upper and lower crust that together are 20 km thick. The rheological model is similar to that illustrated in Fig. 1, except that the transition between the upper and lower crust is more gradual to avoid the complication of the mid-crustal lithosphere in Fig. 1. The flow laws used are from refs 12 and 17. The crust is underlain by a stiff upper mantle that stretches homogeneously throughout the computation. The stiffness of this layer allows it to be modelled as a relatively thin 5 km-thick layer that is allowed to slip freely in the x -direction, but which is constrained to zero vertical deformation at its base. Experiments with grids as deep as 175 km show no differences between the shallow and deep grids, whereas the shallow grid allows much more resolution in the horizontal direction. Freely slipping faults are introduced at 50-km intervals in the upper crust using the recently developed 'slippery-node' method¹⁸. Note that the stress axes in the vicinity of the fault in Fig. 5 are rotated so that one principal stress is perpendicular to the fault plane, as required by the slip condition. The model is extended at a strain rate of $2 \times 10^{-15} \text{ s}^{-1}$, and the stresses that develop near the fault after 26,000 years are plotted in Fig. 5. Only deviatoric stresses are plotted here. Such a plot avoids problems with a mild instability in the finite-element code known as 'pressure checkerboarding' (ref. 19) that develops under the application of large viscous strains. Although

this instability limits the accuracy with which the pressure field can be determined, it does not affect the accuracy of either the deviatoric stresses or the displacements¹⁹. The finite-element grid is much broader than the region shown, minimizing the effect of artificial (velocity) boundary conditions at its ends: the entire grid is 400 km long and 25 km deep and contains 8 faults, 920 elements and 1,032 nodes. The section shown in Fig. 5 focuses on the end of one crustal block.

It is clear from Fig. 5 that stresses are rotated at mid-crustal levels. The greatest ($\sim 45^\circ$) rotation occurs at ~ 6 km depth beneath the hanging wall of the steep segment of the fault, although smaller rotations are apparent even at depths of 4 km. The lower viscosities at depths of 7 km and deeper prevent the development of any significant deviatoric stresses at the time of this plot, although at earlier times 45° stress rotations occurred at deeper levels. Any faults that follow the directions dictated by these stress directions will thus make a sharp turn at the brittle-ductile transition, flattening out in a manner consistent with that inferred for the low-angle normal faults in the Basin and Range province.

Stress and slip on the boundary faults

It may seem surprising that the horizontal stress near the fault in Fig. 5 does not fall to zero. Although the fault surface itself enforces a condition of zero resolved shear stress, and although it cuts completely through the elastic part of the upper crust, a significant horizontal stress σ_{xx}^{ft} nevertheless develops at the fault. The ultimate origin of this stress is in the viscous lower crust. Although the fault slips freely, the geometry of the fault requires that a given horizontal slip must be accompanied by a corresponding vertical slip, as open gaps cannot develop in the Earth. But the upper-crustal block rests on a viscous substratum that must flow from beneath the hanging wall to the footwall to accommodate the vertical displacement (see Fig. 6). This flow requires a stress to drive it, and this stress ultimately appears as an additional horizontal stress σ_{xx}^{ft} which acts in the vicinity of the end of the upper-crustal block. (Note that the development of great topographic differences across the fault may also contribute to this stress. Such stresses are ignored here as being generally smaller than the viscous contribution and entirely negligible if erosion and deposition outpace tectonic uplift.)

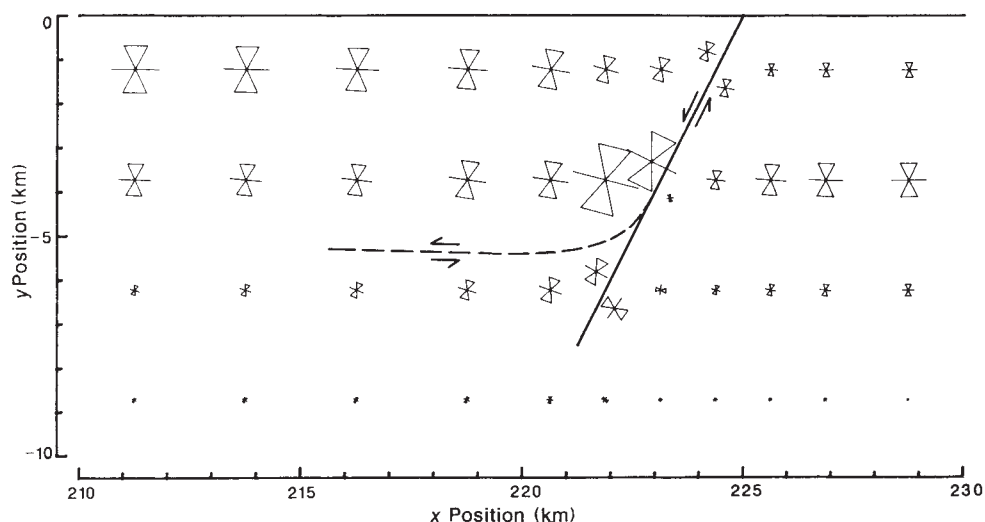
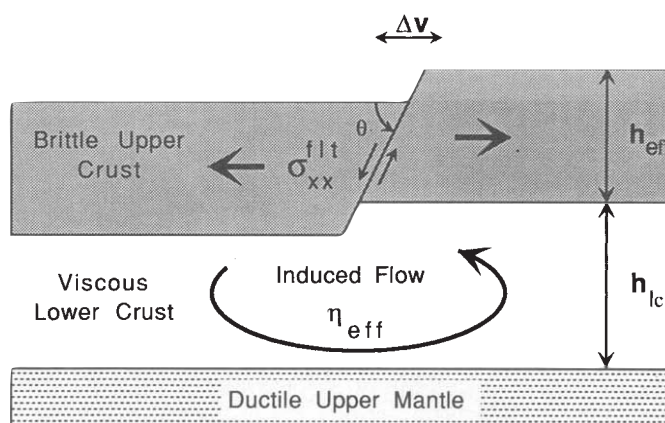


FIG. 5 Deviatoric stresses in a finite-element model of crustal extension 26,000 years after the beginning of extension at a strain rate of $2 \times 10^{-15} \text{ s}^{-1}$. The viscoelastic material obeys a power-law flow relation with $n=3$ and has the following effective viscosities (at the above strain rate): $\eta_{\text{eff}} = 3 \times 10^{24} \text{ Pa s}$ between 0 and 2.5 km depth, $4 \times 10^{22} \text{ Pa s}$ to

5 km, $4 \times 10^{21} \text{ Pa s}$ from 5 to 7.5 km, $1 \times 10^{22} \text{ Pa s}$ from 7.5 to 10 km, $6 \times 10^{20} \text{ Pa s}$ from 10 to 20 km, then 10^{24} Pa s from 20 to 25 km (stiff upper mantle, for which $n=3.5$). The largest symbol represents a stress of $7.35 \times 10^7 \text{ Pa}$. The horizontal dashed line illustrates the trend of a fault forming under the influence of the stress field in the lower crust at this time.

FIG. 6 Schematic illustration of flow in the vicinity of a fault cutting an upper-crustal block and the induced flow that results in a horizontal stress σ_{xx}^{fit} associated with slip on the fault.



A simple geometrical derivation gives the following approximate equation

$$\sigma_{xx}^{\text{fit}} \approx \left(\frac{2\eta_{\text{eff}}}{h_{\text{eff}}} \right) \Delta v \quad (6)$$

In this equation it is remarkable that there is no dependence on either fault dip or on the thickness of the lower crustal layer, a conclusion that is supported by a number of finite-element runs with different fault dips and crustal thickness. It was also verified that equation (6) roughly holds when the crustal block is terminated by a graben rather than by a single normal fault: for a graben σ_{xx}^{fit} is lowered by less than a factor of two.

The existence of the stress σ_{xx}^{fit} at the ends of the otherwise free upper-crustal blocks may play a primary part in determining whether extension takes place by the creation and separation of large crustal blocks, as in the Basin and Range, or whether it is accommodated on many closely spaced normal faults. If σ_{xx}^{fit} is small when compared to the strength of the crust then large relatively rigid blocks can be expected to develop in the extended region, but if σ_{xx}^{fit} is much larger than the crust's strength then it would be expected to break up into many small fragments. In the finite-element simulation of Fig. 5, σ_{xx}^{fit} was $\sim 7 \times 10^7$ Pa, corresponding to a viscosity η_{eff} of 6×10^{21} Pa s. This relatively large stress may preclude the formation of isolated crustal blocks, which would thus require a lower viscosity. Such a lower viscosity may have been realized by a higher thermal gradient in the Basin and Range at the time of the major detachment faulting than is observed today.

Conclusions

The above results from both analytical and finite-element models illustrate the principal conclusion that it is the rheological change between the upper and lower crust that is responsible for the low-angle normal faults observed in the Basin and Range province, not a special or local loading condition. The astute reader will note that the largest stress differences are in the brittle upper crust, not in the viscous lower crust. To explain

the occurrence of low-angle normal faults, we must suppose that the earthquake that accompanies sudden slip on the fault initiates near the base of the brittle crust. Displacement during the event propagates rapidly towards the surface along a steep fault as seen in the 1983 Borah Peak earthquake²⁰. Further down in the viscoelastic lower crust strain propagation may be slower and the direction taken by the propagating rupture is either guided into a horizontal orientation by the ambient stress field after the seismic event or, more likely, may accumulate by aseismic slip for a long time before the event. This may explain the oft-cited observation that, whereas low-angle normal faults seem to be common in the Basin and Range, low-angle earthquake focal mechanisms are not.

The fact that a change in rheology can induce drastic changes in fault orientation is not confined to the Basin and Range problem alone. On a much smaller scale, it is frequently observed that when faults in competent bedded rocks cross horizons of more deformable material, such as shale horizons or evaporite beds, the cross-cutting faults suddenly change direction, being 'refracted' across the deformable bed. This refraction may be seen as simply the result of viscoelastic stress relaxation in the deformable bed, coupled with a small amount of shear displacement between the surrounding competent units. A theory of fault refraction along these lines has already been presented by Bradshaw and Zoback¹⁵.

Similarly, thrust faults are often inferred to 'root' in a horizon of especially deformable rock in which their orientations become parallel to the bedding²¹. This situation is sometimes called 'detachment' faulting in compressional environments or 'detachment' faulting when extension dominates. This situation may again be simply an example of the 45° rotation of stresses that invariably occurs in a sheared viscoelastic layer. Stress (and hence fault) rotation due to a rheologically weak layer is thus a common situation in geology, and the low-angle normal faults observed in the Basin and Range province and in metamorphic core complexes should only be regarded as unusually large examples of this phenomenon. □

Received 4 September; accepted 6 December 1989.

- Proffett, J. M. *J. Geol. Soc. Am. Bull.* **88**, 247–266 (1977).
- Crittenden, M. D. J., Coney, P. J. & Davis, G. H. *Mem. geol. Soc. Am.* **153** (1980).
- Frost, E. G. & Martin, D. L. *Mesozoic–Cenozoic Tectonic Evolution of the Colorado River Region, California, Arizona and Nevada* (Cordilleran, San Diego, 1982).
- Davis, G. A., Anderson, J. L., Frost, E. G. & Shackelford, T. J. *Mem. geol. Soc. Am.* **153**, 79–129 (1980).
- Wernicke, B. *Nature* **291**, 645–648 (1981).
- Lee, J., Miller, E. L. & Sutter, J. F. *Spec. Publ. geol. Soc. Am.* **28**, 267–298 (1987).
- Davis, G. A., Lister, G. S. & Reynolds, S. J. *Geology* **14**, (1986).
- Anderson, E. M. *The Dynamics of Faulting* 1–206 (Oliver and Boyd, Edinburgh, 1951).
- Yin, A. *Tectonics* **8**, 469–482 (1989).
- Spencer, J. E. & Chase, C. G. *J. geophys. Res.* **94**, 1765–1775 (1989).
- Smith, R. B. & Bruhn, R. L. *J. geophys. Res.* **89**, 5733–5762 (1984).
- Kirby, S. H. & Kronenberg, A. K. *Rev. Geophys.* **25**, 1219–1244 (1987).

- McGarr, A. & Gay, N. C. *A. Rev. Earth planet. Sci.* **6**, 405–436 (1978).
- Turcotte, D. L. & Schubert, G. *Geodynamics: Applications of Continuum Physics to Geological Problems* 1–450 (Wiley, New York, 1982).
- Bradshaw, G. A. & Zoback, M. D. *Geology* **16**, 271–274 (1988).
- Melosh, H. J. *Pure appl. Geophys.* **115**, 429–439 (1977).
- Kirby, S. H. & Kronenberg, A. K. *Rev. Geophys.* **25**, 1680–1681 (1987).
- Melosh, H. J. & Williams, C. A. *J. geophys. Res.* **94**, 13961–13973 (1989).
- Hughes, T. J. R. *The Finite Element Method* Ch. 4 (Prentice-Hall, Englewood Cliffs, NJ, 1987).
- Stein, R. S. & Barrientos, S. E. *J. geophys. Res.* **90**, 11,355–11,366 (1985).
- Suppe, J. *Principles of Structural Geology* 1–537 (Prentice-Hall, Englewood Cliffs, NJ, 1985).
- Dziwonski, A. M., Hales, A. L. & Lapwood, E. R. *Phys. Earth planet. Inter.* **10**, 12–48 (1975).

ACKNOWLEDGEMENTS. Greg Radel participated in the early finite-element work on this problem, and I thank Clem Chase and Terry and Jan Tullis for discussions on this approach to low-angle normal faults.

Interleukin-1 receptor antagonist activity of a human interleukin-1 inhibitor

Charles H. Hannum[§], Carol J. Wilcox, William P. Arend^{*}, Fenneke G. Joslin^{*}, David J. Dripps, Patricia L. Heimdahl, Lyman G. Armes, Andreas Sommer[†], Stephen P. Eisenberg & Robert C. Thompson[‡]

Synergen Inc., 1885 33rd Street, Boulder, Colorado 80301, USA

^{*} Division of Rheumatology, Department of Medicine, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

Three interleukin-1 inhibitors have been purified to homogeneity from medium conditioned by human monocytes. Partial sequence analysis and digestion with N-glycanase indicate that these are glycosylation forms of a single protein. The protein binds to the interleukin-1 receptor but has no interleukin-1-like activity, even at very high concentrations, and is therefore a pure receptor antagonist.

THE host response to injury and infection involves the production of interleukin-1 (IL-1). There seem to be two main forms of IL-1: IL-1 alpha with a pI of ~5 and IL-1 beta with a pI of ~7. Although these proteins have limited sequence homology they seem to act identically. Both are synthesized primarily by cells of the monocyte-macrophage lineage in response to various inducers including endotoxin and tumour necrosis factor alpha, and both act on the same target cells through common receptors. The cellular response to IL-1 varies with the identity of the target cell. In endothelial cells, the cytokine induces a pro-coagulant, antifibrinolytic, and adhesive character, in fibroblasts and chondrocytes it induces the synthesis of prostaglandins and proteases, and in T lymphocytes it induces the production of interleukin-2 and the expression of the IL-2 receptor (refs 1-3).

The potent pleiotropic effects of IL-1 on host physiology suggests that its actions must be tightly regulated if they are not to be injurious. Although the best-characterized regulation of IL-1 is at the level of its synthesis, there are reports of IL-1 inhibitors that can also regulate its action after synthesis^{4,5}. So far, none of these inhibitors has been purified and characterized, although an inhibitor from urine that could block the IL-1 receptor has been reported⁶. IL-1 inhibitory activity is induced in medium conditioned by monocytes grown on adherent immune complexes⁷ or IgG⁸. We have now purified three proteins with IL-1-inhibitory activity from this conditioned medium. We show that these represent a nonglycosylated and two glycosylated forms of a previously unidentified polypeptide. These proteins block the action of IL-1 by binding to IL-1 receptors with about the same affinity as IL-1; they have no IL-1-like actions. The mechanism of action of the inhibitor is therefore novel in that a pure receptor antagonist has not been previously reported.

The specificity of this IL-1 antagonist⁸, and the fact that it is produced by the same cells that produce IL-1, support the view that it is a physiologically significant regulator of IL-1 and that an appreciation of its properties will be essential to understanding the function of IL-1.

Purification

IL-1 inhibitor activity is induced by plating monocytes on adherent IgG. We therefore radiolabelled the inhibitor by cultur-

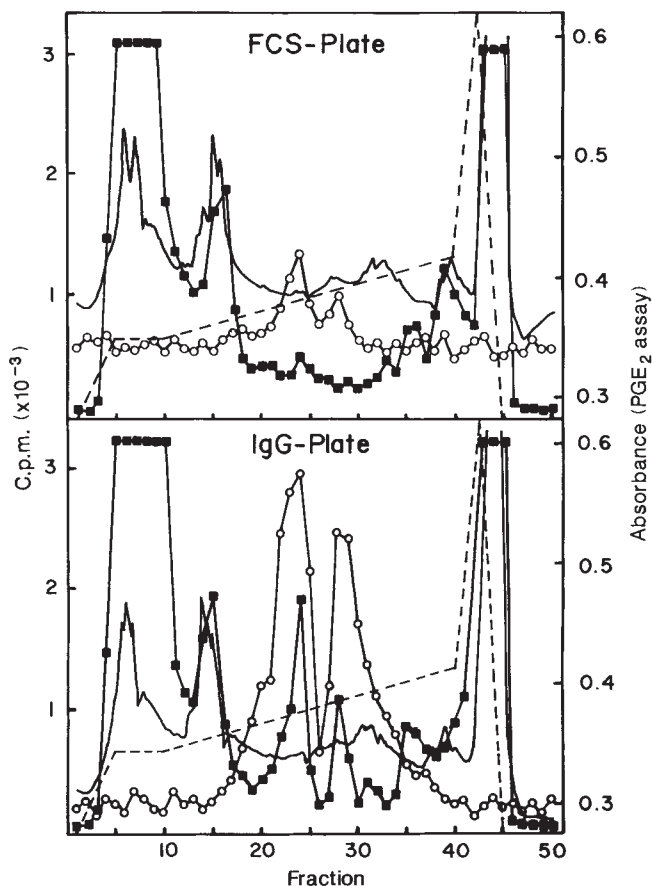


FIG. 1 Mono-Q chromatography of culture supernatants from human monocytes metabolically labelled for 48 h with [³⁵S] methionine in serum-free medium.

METHODS. Human monocytes were collected by leukapheresis from the blood of normal donors and enriched by centrifugation on Lymphoprep^{7,8}. These cells were plated on culture dishes coated with either FCS (upper panel) or human IgG (lower panel) in serum-free RPMI-1640 medium (Mediatech) containing 0.75 µg ml⁻¹ cold methionine (15 µg ml⁻¹ is normal), to which was added 0.5 mCi [³⁵S] methionine (1151 Ci mmol⁻¹) per 10⁷ cells. Crude culture supernatants were made up to a concentrations of 1 mM in phenylmethylsulphonyl fluoride (PMSF), 5 mM in Tris buffer, pH 7.6, and 1.0 M in NaCl, incubated for 1 h on ice and centrifuged at 10,000 r.p.m. for 15 min. The supernatants, which typically retained virtually all of the inhibitory activity, but only ~10% of the total protein found in the crude supernatants, were dialysed against 0.025 M Tris buffer, pH 7.6, containing 0.1% sucrose, the A-buffer for anion exchange chromatography. After dialysis, the solutions were passed through 0.22-micron filters and loaded onto a 1-ml Mono-Q-Superose (Pharmacia FPLC) column. The bound proteins were eluted with a complex NaCl gradient (in A-buffer) from 0-250 mM. Samples of each fraction were tested for radioactivity (■) and IL-1 inhibitory activity (○). The bioactivity assay used was the inhibition of IL-1 beta-induced synthesis of PGE₂ by human foreskin fibroblast cells⁹.

[§] Present address: DNAX, 901 California Avenue, Palo Alto, California 94304-1104, USA.

[†] Present address: Biogen Inc., Richmond, California 94806, USA.

[‡] To whom correspondence should be addressed.

ing the monocytes on plates coated with IgG in medium containing [35 S] methionine. The monocyte-conditioned medium that we used for the purification was supplemented with a small quantity of this radioactive medium to help us identify fractions containing IgG-inducible proteins. Preliminary experiments showed that the inhibitory activity eluted from a Mono-P (chromatofocusing) column as a broad peak from pH 4.5 to 5.3. Therefore we initially fractionated the monocyte-conditioned medium by anion exchange chromatography to look for radioactive species co-eluting with the bioactivity.

Chromatography of the IgG-induced conditioned medium on a Mono-Q column (Fig. 1, lower panel) showed that the inhibitory activity eluted as two or three peaks between 70 and 85 mM NaCl and that each of these peaks contained radioactivity. By contrast, a parallel experiment with medium from monocytes grown on plates coated with fetal calf serum (FCS) showed very little bioactivity or radioactive material at these positions (Fig. 1, upper panel). When we analysed column fractions from the IgG- and FCS-plated monocyte-conditioned media by SDS-PAGE and silver-stained them, only the conditioned medium from IgG-plated cells showed a new protein at positions corresponding to a relative molecular mass of 18,000 (M_r 18K) and two new proteins at positions corresponding to an M_r of 22K (Fig. 2b). Each of these three bands was radioactive, as shown by autoradiography (Fig. 2d). The congruence of the bioactivity and the radioactivity on Mono-Q chromatography, and the correspondence between the M_r s of the induced proteins and a previous estimate of the M_r of the monocyte-derived inhibitor⁷, indicated that all three of the protein bands represent naturally occurring IL-1 inhibitors. We termed these protein inhibitors x (18K), alpha (α ; 22K), and beta (β , 22K), in the order of their elution from the Mono-Q column.

The three species that we identified constituted >90% of the IL-1 inhibitor bioactivity in the monocyte-conditioned media. But prolonged exposure of the autoradiograms showed that at least three other IgG-inducible proteins of M_r 18–22K eluted from the mono-Q column in the same region as x, alpha, and beta proteins; these other proteins could also be IL-1 inhibitors.

We purified each of the three IL-1 inhibitors to homogeneity by consecutive anion exchange, gel filtration, and reverse-phase HPLC; we located the inhibitor-containing fractions first on the basis of radioactivity, and then on the basis of bioactivity. To assess the purity of the inhibitor in these fractions, we analysed each sample by silver-stained SDS-PAGE. The purification was complicated only by each species showing multiple sub-forms on reverse-phase HPLC and the yields being very poor in this final step. Figure 3 shows the overlapping profiles of protein, radioactivity and bioactivity (receptor competition assay) for the final purification step, as well as a silver-stained gel and an autoradiogram of the purified preparations of x, alpha, and beta proteins used for protein analysis. A summary of a typical preparative purification is presented in Table 1.

FIG. 2 Silver-stained 12.5% SDS-PAGE (upper panels)¹⁴ and corresponding autoradiograms (lower panels) of chromatographic fractions of the uninduced (FCS and induced IgG) supernatants shown in Fig. 1. The three species of interest, proteins x (18K), alpha (22K), and beta (22K) are indicated by arrows. The fractions containing the most dense silver-stained (and most radioactive) bands were also the fractions containing the highest levels of inhibitory activity. The radioactive band at the position corresponding to an M_r of 14K is that of human lysozyme.

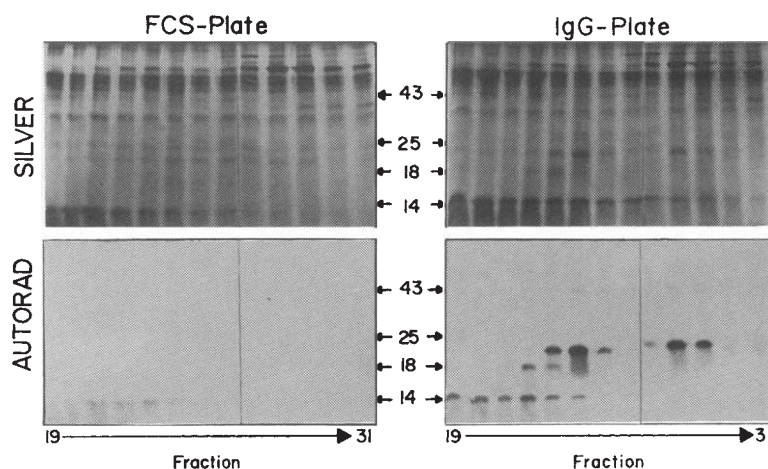


TABLE 1 IL-1ra purification

Sample	Total Protein (mg)	Activity Units	Specific activity (U mg ⁻¹)	Purification (-fold)	Yield (%)
Crude	1,448	28,592	19.7	—	100
QM salt supernatant	111	22,447	202.2	10.3	78.5
Mono Q					
x	0.80	2,414	3,018		
alpha	1.73	6,371	3,682	101.8	51.0
beta	4.74	5,790	1,222		
	7.27	14,575	2,008		
Superose 12					
x	0.047	1,288	27,404		
alpha	0.075	2,775	37,000	1,162	23.0
beta	0.165	2,507	15,194		
	0.287	6,570	22,891		
C4-RPHPLC					
x	0.00111	207	189,486		
alpha	0.00156	372	238,462	10,894	3.0
beta	0.00137	288	210,219		
	0.00404	867	214,603		

A unit of inhibitory activity is defined as the amount that produces the same level of inhibition of the binding of 1 ng 125 I-labelled IL-1 alpha protein to EL4-6.1 cells as 1 ng recombinant IL-1 alpha protein. Protein concentrations were determined by either the bicinchoninic acid (Pierce) or Bradford (BioRad) methods. The concentrations of the pure protein samples were estimated from the areas of the peaks of absorbance at 215 nm corresponding to elutions from the C4 column using a standard purified recombinant protein⁹.

IL-1 inhibitors are novel proteins

We used fractions from the reversed-phase HPLC column that were active in the receptor competition assay and pure as determined by silver-stained SDS-PAGE for sequence analysis. Of ten samples only two yielded significant sequence information. Sequencing of the alpha protein yielded an N terminus of RPSGRKSSKMQAFOXIXDVNQ (single-letter amino-acid code), in which we were unsure about the identification of the N-terminal arginine (R) and could not identify the residues denoted by X. The methionine residue (M) was radioactive, indicating that the molecule was probably the inhibitor. A preparation of the x protein yielded the same sequence, but in this case identification of the N-terminal arginine was unambiguous. Three preparations of the beta protein yielded no N-terminal sequence information.

To obtain further sequence information on the alpha protein, and initial sequence information on the beta protein, we digested these proteins with endoproteinase Lys C and endoproteinase Asp N, respectively. We obtained the following peptide sequences from the alpha protein (single-letter amino-acid code): alpha-1, FYQED; alpha-2,3, SGDETRLQLEAVXITDLSEN; alpha-4, MQAFRIWDVNQK; alpha-5, YLRNNQLVAGYLQGPVNLEEKIDVVP; and alpha-6, RFAFIRSDSG. We

obtained the following peptide sequences from the beta protein: beta-1, DKRFAFIRSDSG; beta-2, RPSGRKSSKMQAFR; beta-3, DEGVMVTKFYFQED; and beta-4, DVNQKTFYLRNNQLVAGYLQGPVNL. Residues that could not be definitively assigned are underlined.

The alpha and beta (and x) proteins are obviously closely related because there is considerable overlap among these few peptides. It is of interest that the beta-2 peptide also has the sequence of the N terminus, indicating that all three forms of the inhibitor should be directly sequenceable.

We constructed composite sequences for the protein using information from all three forms of the inhibitor (Fig. 4), because there is no evidence of differences between the x, alpha, and beta polypeptides. In these sequences we identified residues representing about 60% of the protein⁹. These sequences are not present in any other polypeptides listed in the Protein Information Resource database. Therefore, we believe that this inhibitor is a previously unidentified molecule.

To determine whether the differences between the three forms of this inhibitor are due to glycosylation, we digested samples of purified x, alpha, and beta proteins with N-glycanase, an enzyme that removes both high-mannose and complex N-linked sugars. The products of digestion of both alpha and beta proteins were similar in size to the x protein (M_r 18K), whereas the x protein itself was unaffected by the enzyme (Fig. 5). This result, with the above sequence information, indicates that x, alpha, and beta proteins are the same polypeptides, except that alpha and beta proteins each contain N-linked sugar contributing about 4K to the M_r . To test whether the degree of glycosylation has any effect in the inhibitory action of this inhibitor, samples of x, alpha, and beta proteins from the C4-reversed-phase HPLC column were assayed for bioactivity in the EL4-6.1 receptor competition assay (Fig. 6a). The x, alpha, and beta proteins all contained 2×10^5 units mg^{-1} , where one unit of activity is defined as the inhibition of binding of the ^{125}I -labelled IL-1 alpha protein observed with 1 ng cold IL-1 alpha protein. Therefore, the N-linked sugar of the inhibitor is not required for full activity in the receptor competition assay.

Inhibitor binding to IL-1 receptor

To determine how the monocyte-derived inhibitor blocks the action of IL-1, we first chromatographed the protein with ^{125}I -

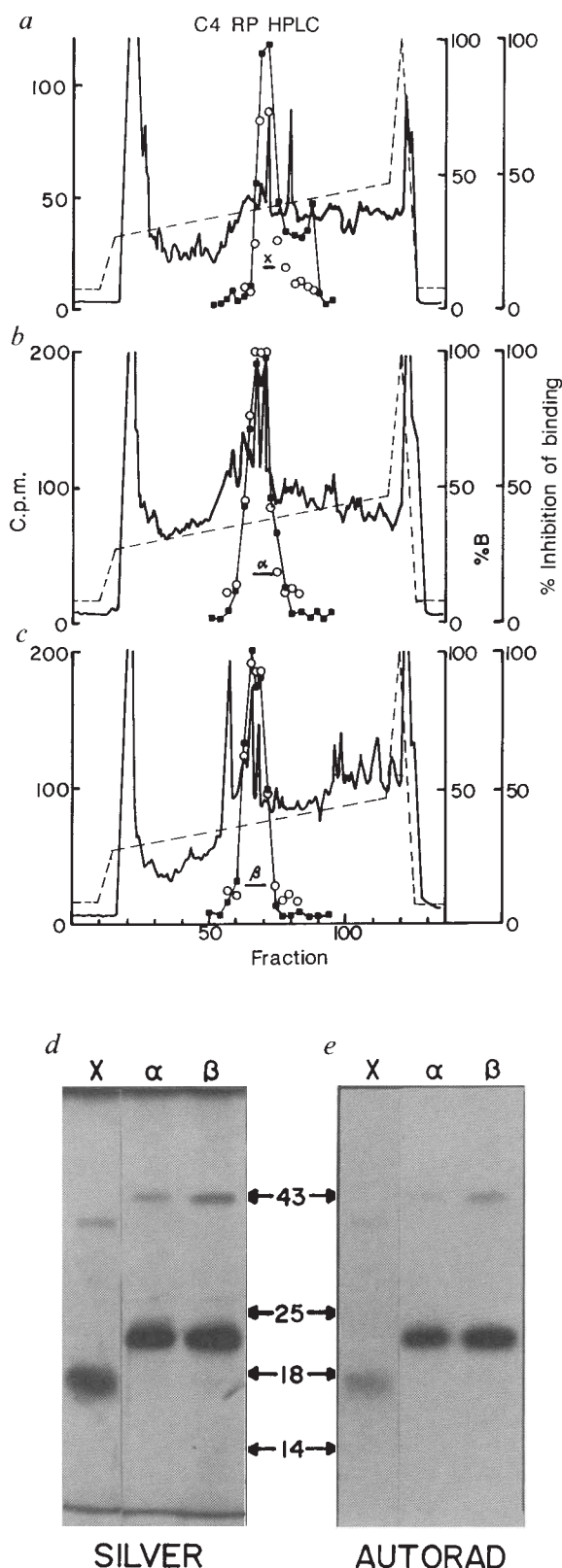
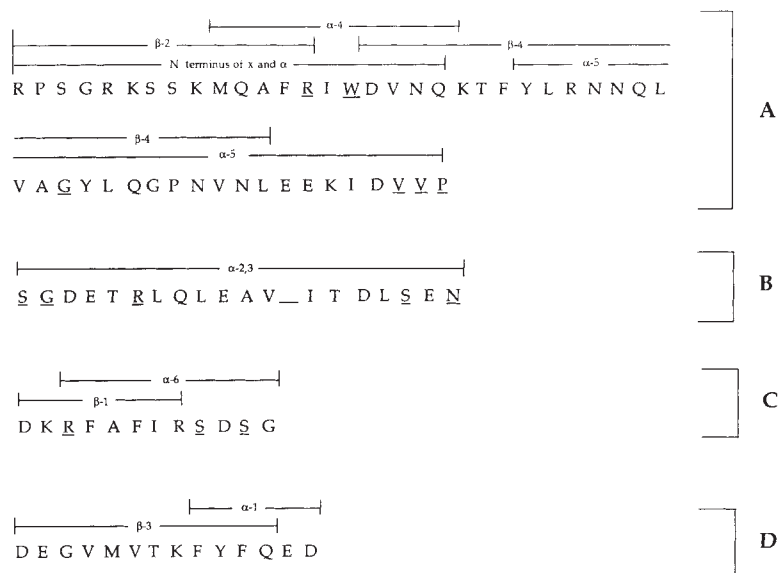


FIG. 3 Reversed-phase (RP) HPLC of proteins x (a), alpha (b) and beta (c) on a Synchrom C4 column. The eluted materials that were collected and used for all subsequent analyses are indicated by the labelled bars. Silver-stained SDS-PAGE (d) and the corresponding autoradiogram (e) of the homogeneously purified proteins x, alpha, and beta, are also shown.

METHODS. Crude monocyte supernatants (typically 500–800 ml per preparation), that were routinely supplemented with metabolically labelled (^{35}S) Met) material representing 5% of the total volume, were chromatographed on a Mono-Q column with a bed volume of 8 ml. Three pools were made, each containing one of the three main forms of the inhibitor. These pools were dried by vacuum centrifugation, resuspended in a minimal volume of 0.025 M Tris buffer, pH 7.6 containing 0.1% sucrose and 0.15 M NaCl, and individually chromatographed at 0.2 ml min^{-1} on a Superose-12 (Pharmacia FPLC) gel filtration column equilibrated in the same buffer. From each column the fractions containing the inhibitor were pooled and chromatographed on a Synchrom C4 reverse-phase column ($4.6 \times 200 \text{ mm}$) using an $\text{H}_2\text{O}/0.1\%$ trifluoroacetic acid $\text{CH}_3\text{CN}/0.1\%$ TFA gradient at 0.2% per min from 25 to 40% $\text{CH}_3\text{CN}/0.1\%$ TFA. In all three cases the inhibitor resolved into several peaks of radioactivity (■) and bioactivity (○) which were always analysed individually. For bioassays, samples were collected in tubes containing $10 \mu\text{l}$ Tris buffer (2 M) pH 7.5. The bioassay used was an IL-1 binding assay in which $1 \text{ ng } ^{125}\text{I}$ -labelled IL-1 alpha protein (labelled using chloramine T^{15}) was reacted with 1×10^6 EL4-6.1 cells for 3 h at 4°C in a binding buffer consisting of RPMI 1640, 0.1% gelatin, 0.2% NaN_3 . Bound counts were separated from unbound on Millititer SV96 plates (Millipore). The solid line shows the optical density at 280 nm, whereas the dashed line represents the gradient. The light bands near the position corresponding to an M_r of 40K in d and e represent dimers of each form of the inhibitor.

FIG. 4 Four composite amino-acid sequences (A, B, C, and D) constructed from the N-terminal data obtained from intact α and β proteins as well as from peptides of α and β . Underlined residues could not be definitely assigned. **METHODS.** Generation and purification of peptides: Homogeneously purified species of α and β proteins were collected by hand off the C4 reversed phase column into siliconized glass tubes, and to each was added 25 μ l Tween-20 (0.2%). The inhibitor-containing fractions were reduced in volume to 50 μ l by vacuum centrifugation, brought to 300 μ l by the addition of 1.0% NH_4HCO_3 , followed by the addition of 1 μ g endoproteinase (1:1 by weight). For the α protein, the enzyme used was Endoproteinase Lys C (Boehringer), whereas the β protein was cleaved with Endoproteinase Asp N (Boehringer). Digestion was at 37 °C for 16 h, and the volume of the reaction mixture was reduced by 50 μ l by vacuum centrifugation. For the α protein, the sample was directly chromatographed, whereas the β -protein sample was first reduced by the addition of 5 μ l of 50 mM DTT in 2 M Tris buffer, pH 8.0, reacted for 30 min at 37 °C, and then carboxymethylated by addition of 1.1 μ mol ^3H -labelled iodoacetic acid in 10 μ l ethanol (reacted for 30 min at 37 °C in the dark). Peptides were separated on a 2.1 $\times$ 250 mm Brownlee Aquapore RP-300 (C8) narrow-bore column at a flow rate of 100 μ l min^{-1} using a Beckman HPLC column fitted with microbore hardware and microbore-compatible pumps. A 200-min 0–100% linear gradient was used ($\text{H}_2\text{O}/0.1\%$ TFA– $\text{CH}_3\text{CN}/0.1\%$ TFA). Microsequence analyses were performed with an Applied Biosystems gas phase sequencer (Model 470) and an on-line phenylthiohydantoin amino-acid analy-



ser (Model 120) using the 03RPTH sequence program and the manufacturer's recommended program and solvents for the analyser. Sequencer analyses were carried out on 5–50 pmol sample, and repetitive yields of 87–94% were obtained. Analyses resulted in a useful signal of 2–15 pmol.

labelled IL-1 on a gel-filtration column. We did not detect the formation of a complex between these species (data not shown), indicating that the inhibitor does not block the receptor interaction by binding the cytokine. By contrast, incubation of ^{35}S -labelled inhibitor with EL4-6.1 cells showed that the molecule forms a strong complex with the IL-1 receptor.

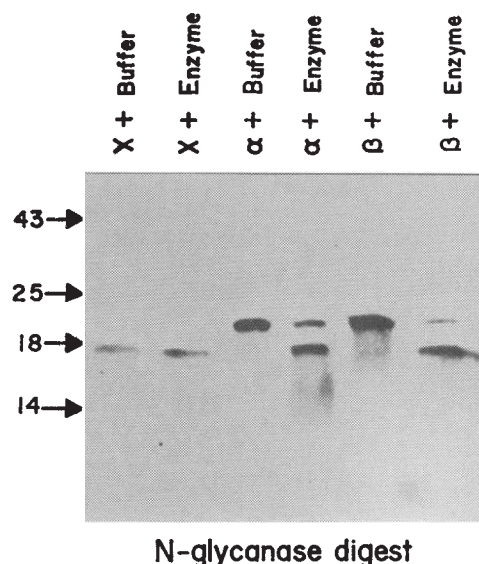
For this experiment we used recombinant inhibitor prepared in *Escherichia coli* using a complementary DNA clone derived from human monocyte RNA as described by Eisenberg *et al.*⁹ The recombinant protein was as effective as the monocyte-derived protein in blocking the binding of ^{125}I -labelled IL-1 α protein to EL4-6.1 cells (Fig. 6a). We prepared highly radioactive recombinant inhibitor by growing the *E. coli* cells on M9 medium containing [^{35}S] sulphate and purified by chromatography on a Mono-S column. By incubating this material with EL4-6.1 cells under conditions similar to those used in the IL-1 binding assay, we showed that the radioactive protein binds to the cells (Fig. 6b). The amount of ^{35}S -labelled

inhibitor bound was equivalent to the amount of ^{125}I -labelled IL-1 α protein bound to identical cells in a parallel experiment. The binding of the inhibitor was blocked when we included excess unlabelled IL-1 α or IL-1 β in the incubation mixture, indicating that the inhibitor binds specifically to the IL-1 receptor. Binding of the ^{35}S -labelled inhibitor was also blocked by excess unlabelled inhibitor, indicating that the binding is saturable. At all concentrations the extent of inhibition of the binding of the ^{35}S -labelled inhibitor by cold inhibitor and by IL-1 α and IL-1 β proteins was similar, indicating that the antagonist and the agonists have similar affinities for the IL-1 receptor on EL4-6.1 cells.

Protein ligands that are pure receptor antagonists, that is, ligands that show no agonist activity even at very high concentrations, have not been reported previously. Consequently, it was important to determine whether this inhibitor has any agonist activity. To do this we incubated human dermal fibroblasts with 10 $\mu\text{g ml}^{-1}$ recombinant inhibitor and measured the

FIG. 5 SDS-PAGE analysis of inhibitor samples that were digested with the deamidase N-glycanase (Genzyme), an enzyme that removes all N-linked glycosylations.

METHODS. Protein samples (200 ng) were reacted for 18 h at 37 °C with either 0.3 U N-glycanase or buffer alone according to the manufacturer's specifications.



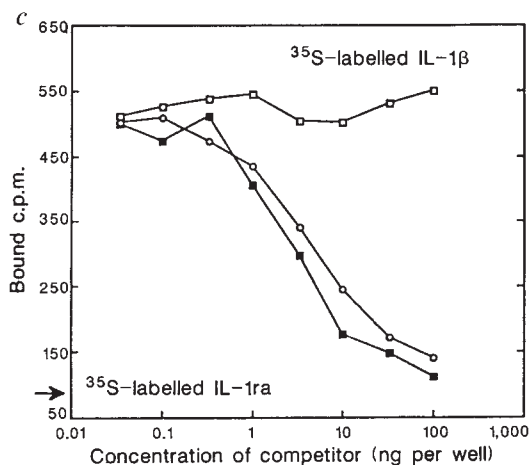
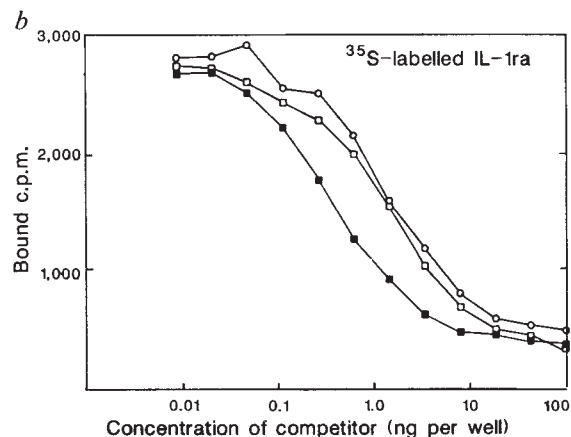
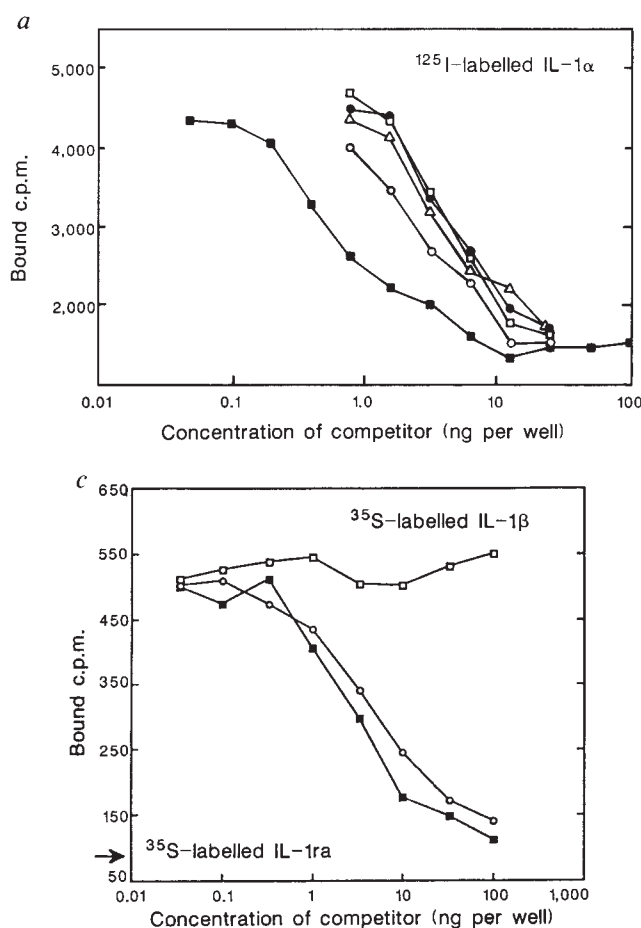


FIG. 6 a, Receptor-competition assays in which ~ 1 ng ^{125}I -labelled IL-1 α was inhibited from binding to 1×10^6 EL4-6.1 cells by varying amounts of cold IL-1 α (■), antagonist-form α (●), antagonist-form β (□), antagonist-form γ (△), and recombinant antagonist (○). The concentrations of the purified forms of antagonist were estimated from the areas of the peaks of absorbance at 250 nm eluting from the C4 column. b, Receptor-competition assay in which ~ 1 ng ^{35}S -labelled antagonist was inhibited from binding to EL4-6.1 cells by varying amounts of cold recombinant IL-1 α (■), IL-1 β (○), and recombinant antagonist (□). c, Receptor-competition assay in which 1 ng ^{35}S -labelled IL-1 β was inhibited from binding to 1×10^6 70Z/3 cells by varying amounts of unlabelled IL-1 α (■), IL-1 β (○), and recombinant antagonist (□). The arrow indicates the CPM bound to 70Z/3 cells that were treated with 1 ng ^{35}S -labelled IL-1ra.

prostaglandin E_2 (PGE_2)⁹. The measured levels of PGE_2 were 1.1 ng ml^{-1} in the supernatants from inhibitor-treated cells and 1.0 ng ml^{-1} in cells that were not exposed to the inhibitor. By contrast, cells exposed to 0.3 ng ml^{-1} IL-1 beta protein synthesized 27.2 ng ml^{-1} PGE_2 in a parallel experiment. Because this inhibitor specifically competes with IL-1 for binding to its receptor, but has no measurable agonist activity, we termed this molecule interleukin-1 receptor antagonist (IL-1ra).

Recent experiments have indicated that there are more than one type of IL-1 receptor in the mouse^{10,11}. Consistent with this, we find that the murine pre-B cell line 70Z/3 binds IL-1 alpha and beta, but does not seem to bind ^{35}S -labelled IL-1ra above background levels. Additionally, unlabelled IL-1ra does not block the binding of ^{35}S -labelled IL-1 beta to these cells (Fig. 6c).

Discussion

We have purified and partially sequenced three human monocyte IL-1 receptor antagonists and shown them to be glycosylation variants of a single protein. We guided the purification process by following radioactivity, introduced into the protein by metabolically labelling during IgG-induction, and by measuring bioactivity. The antagonist (IL-1ra) blocks IL-1 action by binding to the IL-1 receptor but shows no IL-1-like activity in the

fibroblast PGE_2 -synthesis assay. It seems that IL-1ra represents a novel species, that is, a protein receptor ligand that is a pure antagonist.

The function of IL-1ra remains to be determined. Obviously any naturally occurring molecule that can specifically inhibit a cytokine as pleiotropic as IL-1 could have several roles. In physiological and pathophysiological conditions where IL-1 is a significant effector it is possible that IL-1ra is also present, acting to maintain or restore a steady state by modulating IL-1 action. Measurements of the relative levels of IL-1 and IL-1ra in serum and, perhaps more importantly, in compartments such as arthritic synovia, should give a clearer picture of how these molecules act together to maintain a physiological steady state. To this end, we are now developing an enzyme-linked immunosorbent assay (ELISA) immunosorbent assay for IL-1ra using monoclonal antibodies.

On the basis of their similar sizes and charges it is quite likely that IL-1ra is the same protein as the inhibitor of M_r 23K that has been partially purified from the urine of febrile patients and those with myelomonocytic leukaemias^{6,12,13}. The urine-derived protein is also thought to prevent the binding of ^{125}I -labelled IL-1 alpha to EL4-6.1 cells because of its ability to block IL-1 binding when pre-incubated with the cells⁶. □

Received 1 December; accepted 19 December 1989.

- Oppenheim, J. J., Kovacs, E. J., Matsushima, K. & Durum, S. K. *Immun. Today* **7**, 45-56 (1986).
- Le, J. & Vilcek, J. *Lab. Invest.* **56**, 234-248 (1987).
- Dinarello, C. A. *FASEB J.* **2**, 108-115 (1988).
- Larrick, J. W. *Immun. Today* **10**, 61-66 (1989).
- Dayer, J.-M. & Seckinger, P. in *Interleukin 1-Inflammation and Diseases* (eds Bormford, R. H. R. & Henderson, B.) (Elsevier, in the press).
- Seckinger, P. et al. *J. Immun.* **139**, 1546-1549 (1987).
- Arend, W. P., Joslin, F. G. & Massoni, R. J. *J. Immun.* **134**, 3868-3875 (1985).
- Arend, W. P., Joslin, F. G., Thompson, R. C. & Hannum, C. H. *J. Immun.* **143**, 1851-1856 (1989).
- Eisenberg, S. P. et al. *Nature* **343**, 341-346 (1990).
- Chizzonite, R. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8029-8033 (1989).

- Bomsztyk, K. et al. *Proc. natn. Acad. Sci. U.S.A.* **86**, 8034-8038 (1989).
- Balavoine, J.-F. et al. *J. clin. Invest.* **78**, 1120-1124 (1986).
- Seckinger, P. et al. *J. Immun.* **139**, 1541-1545 (1987).
- Laemmli, U. *Nature* **227**, 680-685 (1970).
- Hunter, W. M. & Greenwood, F. C. *Nature* **194**, 495-496 (1962).

ACKNOWLEDGEMENTS. We thank David Hirsh for his critical evaluation of the manuscript, John Childs for providing us with ^{35}S -labelled IL-1 beta and Pat Kilian of Hoffmann-La Roche, Nutley, New Jersey, for the gifts of unlabelled and ^{125}I -labelled recombinant IL-1 alpha, and 70Z/3 cells. We also thank Evie Verderber and Paula Larsen for assistance in making and assaying the large-scale monocyte supernatants, Laura McGee and Dhyan Atkinson for their help in the preparation of the manuscript, and Tim Klein and Jerry Kuner for their help with the illustrations.

Primary structure and functional expression from complementary DNA of a human interleukin-1 receptor antagonist

Stephen P. Eisenberg, Ron J. Evans, William P. Arend*, Evie Verderber, Michael T. Brewer, Charles H. Hannum† & Robert C. Thompson

Synergen Inc., 1885 33rd Street, Boulder, Colorado 80301, USA

* Division of Rheumatology, Department of Medicine, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

† Present address: DNAX 901 California Avenue, Palo Alto, California 94304-1104, USA

Human monocytes induced with adherent IgG secrete an interleukin-1 receptor antagonist which could be important for the *in vivo* regulation of IL-1 activity. A complementary DNA for this molecule has been isolated from a human monocyte library. Analysis of monocyte RNA indicates that the gene is transcriptionally regulated. The sequence of the receptor antagonist indicates that it is structurally similar to IL-1 β . Expression of the cDNA in *Escherichia coli* yields IL-1 receptor antagonist activity.

INTERLEUKIN-1 (IL-1) is a cytokine¹ produced primarily by mononuclear phagocytes. There are two forms of IL-1, and although they have only slightly homologous primary structures², they have similar effects on a variety of target cells³. They both stimulate the production of prostaglandin and collagenase by fibroblasts, induce the synthesis of interleukin-2 by T cells, and induce adhesion proteins and a procoagulant-antifibrinolytic character in vascular endothelial cells. On the basis of these activities and the increased serum levels of IL-1 seen in several disease states⁴, IL-1 is implicated in a range of inflammatory diseases, such as rheumatoid arthritis. Also, injections of IL-1 in rabbit knee joints causes swelling and the loss of proteoglycan⁵, and intravenous injections of IL-1 into rabbits lowers blood pressure and dramatically reduces the number of circulating white blood cells⁶.

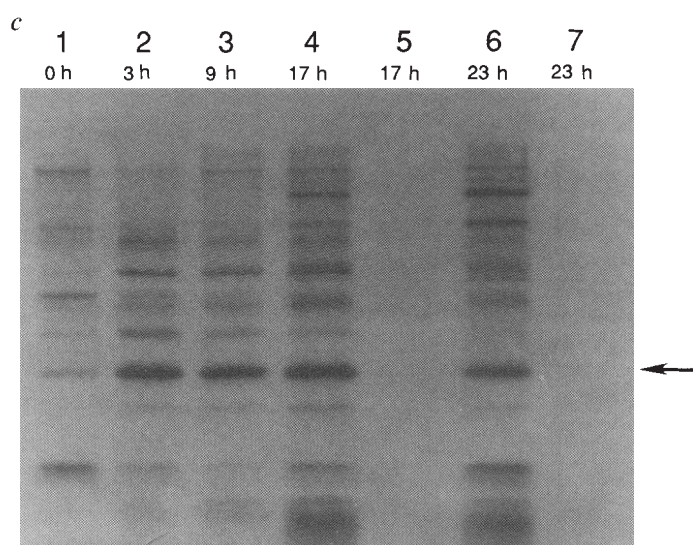
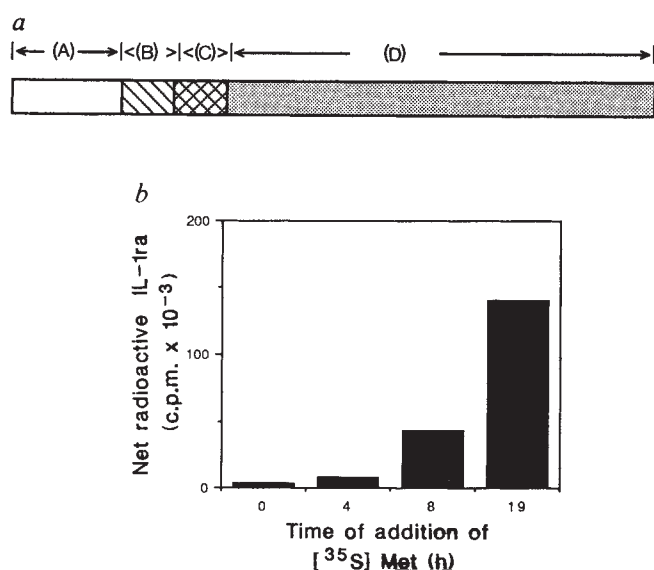


FIG. 1 Analysis of pulse-labelled proteins from IgG-stimulated monocytes. *a*, Design of labelling protocol. Mononuclear leukocytes (which include monocytes and lymphocytes) were isolated from normal donor blood by Lymphoprep gradient centrifugation. These cells were plated on IgG-coated culture dishes at a density of 6.5×10^6 cells ml^{-1} in RPMI medium containing $\frac{1}{50}$ $\times$ normal concentration of methionine ('low Met RPMI')¹² and incubated at 37 °C in 5% CO₂. The schedule for metabolically labelling the cells is indicated: (A) 0–x hours, prelabelling period; (B) 2-h pulse labelling, [³⁵S]methionine (50 $\mu\text{Ci ml}^{-1}$) added to the medium; (C) 2-h chase, excess cold methionine (100-fold) added to the medium, medium is collected at the end of (C) and is replaced with fresh RPMI medium; and (D) (20–x) hour recovery period, medium is collected at the end of this period. *b*, [³⁵S]IL-1 receptor antagonist identified by column chromatography. Supernatants taken after periods (C) or (D) were applied to a Mono-Q column (Pharmacia) and eluted using a NaCl gradient from 0 to 0.1 M, in 25 mM Tris-HCl buffer, pH 7.6, 0.1% sucrose. Fractions were collected and the amount of radio-

activity was determined by liquid scintillation counting. The times at which labelling began (beginning of (B)) and the net radioactivity (total – background (6,000 c.p.m. per fraction)) in the IL-1ra peak¹² from each of the labelled samples collected after period (C) are indicated. In each case, supernatants collected after period (D) contained no detectable [³⁵S]IL-1 antagonist, indicating that the 2-h chase was effective. *c*, Polyacrylamide gel electrophoresis of [³⁵S]-labelled supernatants from pulse-labelled monocytes. Supernatants were collected after periods (C) and (D), and 25 μl of each sample was loaded on a 15% SDS-polyacrylamide gel³³. After electrophoresis, the gel was fixed, treated with Enhance (NEN), dried and exposed to Kodak XAR film to produce the fluorogram shown. The times at which labelling began are indicated at the top of each lane. Lanes 1, 2, 3, 4 and 6 are samples taken after period (C); lanes 5 and 7 are samples taken after recovery period (D). The arrow indicates the putative IL-1ra band at a position corresponding to an *M_r* of 22K. The absence of radioactivity in lanes 5 and 7 indicates that the cold methionine chase was effective.

The potent pleiotropic inflammatory effects of IL-1 are likely to be stringently controlled *in vivo*. There are many ways in which this regulation could be achieved. One of the more interesting mechanisms would be through the synthesis of a protein inhibitor of IL-1 that could regulate IL-1 action after its synthesis. In addition to its potential physiological importance, this inhibitor could be used in the appropriate animal models of human disease to determine whether IL-1 is pathophysiologically significant in human disease. An IL-1 inhibitor could also be useful as a therapeutic agent for treatment of inflammatory diseases.

Several IL-1 inhibitory activities have been described^{7,8}. Most of these seem to inhibit the actions of IL-1 only in the thymocyte proliferation assay⁸. Only two IL-1 inhibitors have been shown

to inhibit IL-1 action in all of the commonly used cellular assays^{9,10}, suggesting that they act directly on IL-1 or on its receptor. We decided to clone the monocyte-derived inhibitor, because the cell that synthesizes this protein is known, and the conditions necessary to induce it have been described^{9,11}. In the accompanying article, Hannum *et al.*¹² describe the purification and properties of a monocyte-derived IL-1 inhibitor that acts as a receptor antagonist (termed IL-1ra) and describe peptide sequences that we used for the synthesis of oligonucleotide probes.

Isolating a cDNA clone for IL-1ra

We performed a preliminary experiment to estimate the levels of functional IL-1ra messenger RNA in monocytes at different

<i>a</i>	Peptide sequence	RNA coding sequence	Oligodeoxyribonucleotide probe sequence (probe no.)
	KMQAF	5'--AARAUGCARGCNUUY-3'	5'-RAANGCYTGCATYTT-3' (1)
	KFYFQED	5'--AARUUYUAYUYCARGARGA-3'	5'-TCYTCYTGRAARTARAAYTT-3 (2)
	MVTKFYF	5'-- AUGGUNACNAARUUYUAYUU-3'	5'-AARTARAAYTTNGTNACCAT-3' (3)
	DVNQKT	5'- GAYGUNAAACARAARAC-3'	5'- GTYTTYTGRTTNACRTC-3' (4)
	NQKTFY	5'-- AAYCARAARACNUUYUA-3'	5'- TARAANGTYTTYTGRTT-3' (5)

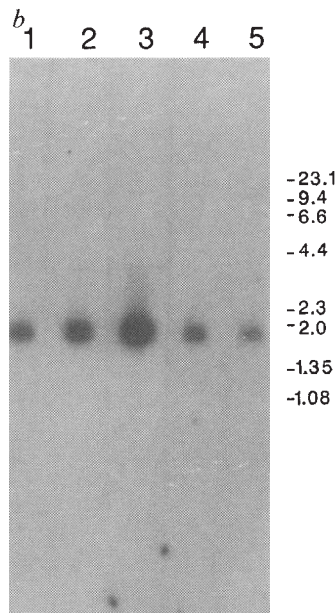


FIG. 2 Sequence of the IL-1ra cDNA. a, Design of oligonucleotide probes. Five degenerate antisense probes were synthesized based on peptide sequences (single-letter amino-acid code)¹² as indicated. Oligonucleotides were synthesized on an Applied Biosystems Model 380A DNA synthesizer and labelled with [γ -³²P]ATP and polynucleotide kinase. In the nucleic-acid sequences shown, R represents A + G, Y represents C + T (or C + U), and N represents all four bases. b, Analysis of cDNA clone IL-1ra-2a by Southern blot. RNA was prepared from human monocytes induced for 17 h as described³⁴. Poly(A)⁺ RNA was prepared by standard techniques using oligo(dT) cellulose³⁵ from Bethesda Research Laboratories, and a λ gt10 library was constructed³⁶ using a cDNA synthesis kit (Boehringer)³⁷ and a λ gt10 cloning kit (Stratagene). *Eco*R1 linkers were purchased from New England Biolabs. The resulting library contained 1.4×10^7 recombinant phage as determined by their ability to form plaques on the *E. coli* strain C600 Hfl. Screening with oligonucleotide probes was performed as described³⁸. The λ gt10 DNA from cDNA clone IL-1ra-2a (see text) was prepared using Lambdasorb (Promega), digested with *Eco*R1, divided into five aliquots, loaded into 5 separate wells and electrophoresed on a 1% agarose gel. The digested DNA was resolved into the two λ arms (32.7 kb and 10.6 kb) and the insert (1.8 kb). The DNA on the gel was blotted onto nitrocellulose³⁹, which was

c

1 GAATTCGGGCTGCAGTCACAGA 23

ATG GAA ATC TGC AGA GGC CTC CGC AGT CAC CTA ATC ACT CTC CTC CTC TTC CTG TTC CAT 83
M E I C R G L R S H L I T L L L F L F H

TCA GAG ACG ATC TCC CGA CCC TCT GGG AGA AAA TCC AGC AAG ATG CAA GCC TTC AGA ATC 143
S E T I C R P S G R K S S K M Q A F R I₁₅

TGG GAT GTT AAC CAG AAG ACC TTC TAT CTG AGG AAC AAC CAA CTA GTT GCT GGA TAC TTG 203
W D V N Q K T F Y L R N N Q L V A G Y A₁₅

CAA GGA CCA AAT GTC AAT TTA GAA GAA AAG ATA GAT GTG GTA CCC ATT GAG CCT CAT GCT 263
Q G P N V N L E E K I D V V P I E P H A₅₅

CTG TTC TTG GGA ATC CAT GGA GGG AAG ATG TGC CTC TCC TGT GTC AAG TCT GGT GAT GAG 323
L F L G I H G G G K L S L V K S D E₇₅

ACC AGA CTC CAG CTG GAG GCA GTT AAC ATC ACT GAC CTG AGC GAG AAC AGA AAG CAG GAC 383
T R L Q L E A V N I T D L S E N R K Q D₉₅

AAG CGC TTC GCC TTC ATC CGC TCA GAC AGT GGC CCC ACC ACC AGT TTT GAG TCT GCC GCC 443
K R F A F I R S D S G P T T S F G S D A₁₁₅

TGC CCC GGT TGG TTC CTC TGC ACA GCG ATG GAA GCT GAC CAG CCC GTC AGC CTC ACC AAT 503
L P G W F L C T A M E A D Q P V S L T N₁₃₅

ATG CCT GAC GAA GGC GTC ATG GTC ACC AAA TTC TAC TTC CAG GAG GAG GAG TAG 567
M P D E G V M V T K F Y F Q E D E₁₅₂

TACTGCCAGGCGCTGCTGTTCCTTCCTGATGCAAGGACTGCAGGACTGCCAGTCCCCCTGCCAGGCTCCCG 637

GCTATGGGGCACTGAGGACGAGCATTGAGGGGTGGACCTCAGAGGCGTCACAAACCTGCTCAGAGACTCTGCC 717

TCCTCTTCACTGACCAAGCCTCATGCTGCTCCAGAAATGCTCTTCTAATGTGTGAATCAGAGCAGCAGCCCTGCA 797

CAAGGCCCTTCCATGTCCGCTCTGCATTGAGGATCAACCCGACCACTGCCCAACCTCTCTCTTGCACCTGCCT 877

CTTCTCTCCATTCACCTTCCATGCTGCTGATTCATGAGGACCTTGATGACCCCAACCAAGTGGCTCCGACACCC 957

TGTTTTTCAAAAAAGAAAGACCACTCCATGAGGAGGTTTTTAAGGGTTTGTGGAAATGAAATATAGGATTTTCATGAT 1037

TTTTTTTTTTCAGTCCCCGTGAAGGAGGCGCTTCATTGGAGATTATGTTCTTTCGGGAGAGGCTGAGGACTTAAAT 1117

ATTCCTGCAATTTGTGAATGATGGTGAAAGTAAGTGTAGCTTTTCCCTCTCTTTTCTCTTTTGTGATGTCGCAAC 1197

TGTGAATAATTAAGATTTATGTTAGTCTATGTTAGCCCAATAATTTTTTTCCTTTTAAACACTTCATATCTGAGACT 1277

CCTCTGTCCAGGCACTGCTGCCAGCCTCCAAAGCTCCATCTCCACTCCAGATTTTTTACAGCTGCTGCACTACTTAC 1357

TCCTATCAGAAATTTCTCAGCTCCCAAGGCTCTGAGCAATGTGGCTCTGAGGGGTTCTTCTCTCTGCTGAGGAAT 1437

AAATGCTCTTACATTTGTAGAGCTTCTGGCACTGGAGACTTGATGAAGATGGCTGCTGCTGCTGCTGCTGCTGCT 1517

ACCAGGCTGGGAGCTCTGAGAGCAGGAAACATGACTGCTATATGTCTCAGGTCCCTGAGGGCCAGGACCTAGCCTCG 1597

CTCTTGGCAGGTACTCAGCAATGAATGCTGTATATTTGGGGTCAAGTTCCTACTCTGCTGACTCAGCTCTGTTT 1677

TACAATAAATCTTGAATGCCAAAAAAGAAAAAAGCCGAATCCCGCCGGAATTC 1740

cut lengthwise into five strips, each representing a different lane of the gel. Each of the strips was hybridized to a different oligonucleotide probe (Fig. 2a, above). Probes are indicated on the top of each strip. After hybridization and washing³⁸, the strips were aligned next to each other and exposed to X-ray film. DNA size markers (in kb) are shown on the right. c, Sequence of the cDNA for IL-1ra. The DNA sequence and the predicted protein sequence of the large ORF are shown. Both strands of the cDNA were sequenced at least once, and in the coding region, both strands were sequenced at least three times^{40,41}. The cDNA sequence and the exon sequences of a genomic clone described elsewhere (S.P.E. *et al.*, manuscript in preparation) are identical except for nucleotide residue 100 (underlined), which is a C in the cDNA and a G in the genomic clone. We believe that the correct residue is a G, because the codon CGA codes for arginine, which agrees with the known N-terminal sequence of the IL-1 antagonist¹². The C residue is apparently due to an error made during the reverse transcription of the IL-1ra mRNA. The sequence of the mature protein, after removal of the 25 amino-acid signal sequence, is shown in bold. Cysteine residues are underlined and the consensus N-linked glucosylation site is indicated by the asterisk. The *Eco*R1 sites and flanking linker sequences are in italics.

times after plating the cells on IgG. We reasoned that the level of active mRNA should be directly related to the rate of synthesis of the protein, which could be measured by determining the amount of radioactive IL-1ra produced during a pulse with [35 S]-methionine. Because the protein is secreted, we used a cold methionine 'chase' so that the pulse-labelled protein could accumulate in the medium (Fig. 1a). We assessed the amount of radioactive IL-1ra made by induced monocytes during the pulse either by ion-exchange chromatography¹² or by estimating the amount of radioactivity in a protein of relative molecular mass 22,000 (M_r 22K) by SDS-PAGE. The results of the initial chromatographic analysis indicated that the level of IL-1ra mRNA is elevated about 20 h after induction (Fig. 1b). Further experiments indicated that the mRNA level is maximal at 15–17 h (data not shown). By contrast, the fluorogram of the gel (Fig. 1c) indicated that the mRNA level increases rapidly to a maximum 3–5 h after induction and remains at this level for the next 14 h. Despite indicating different kinetics of mRNA accumulation, both methods showed that the mRNA was at a maximum at 17 h. We therefore used RNA from 17 h cells to establish a cDNA library in λ gt10.

We constructed the cDNA library by standard methods and screened it with a 64-fold degenerate 15-residue (15 mer) oligonucleotide (probe 1) based on the peptide sequence Lys-Met-Gln-Ala-Phe (KMQAF in single-letter (Fig. 2a)). We re-screened several positive plaques with probe 1 and a second oligonucleotide (a 128-fold degenerate 20mer) based on the amino-acid sequence Lys-Phe-Tyr-Phe-Gln-Glu-Asp (KFYFQED). One clone, IL-1ra-2a, hybridized to both probes.

Digestion of DNA isolated from this clone with *Eco*RI released a 1.8-kilobase (kb) insert that hybridized on a Southern blot to the two probes used in the screening and to three new oligonucleotide probes derived from additional peptide sequence (Fig. 2b). We cloned the 1.8-kb fragment into the *Eco*RI site of M13mp19¹³ and determined its nucleotide sequence.

The cDNA (Fig. 2c) has a single open reading frame (ORF) coding for 177 amino acids that is preceded by a relatively short 5' untranslated region (UTR) of 14 nucleotides (excluding the *Eco*RI linker sequence) and is followed by a long 3' UTR of 1,133 nucleotides excluding the poly(A) tail. The clone is close in length to that of the IL-1ra mRNA (~1.8 kb) estimated by northern-blot analysis (Fig. 3). On the basis of primer extension analysis, the 5' end of the cDNA is, at most, five nucleotides from the 5' end of the transcript (data not shown).

The ORF of the clone starts at an ATG codon flanked by

TABLE 1 Inhibition of IL-1-induced PGE₂ synthesis by IL-1ra

Lysate	PGE ₂ (ng ml ⁻¹)		
	*1:1	*1:10 ²	*1:10 ⁴
pT5T	23 ($\pm$ 3)	>63	>63
pRJ2	8 ($\pm$ 2)	5 ($\pm$ 1)	23 ($\pm$ 4)

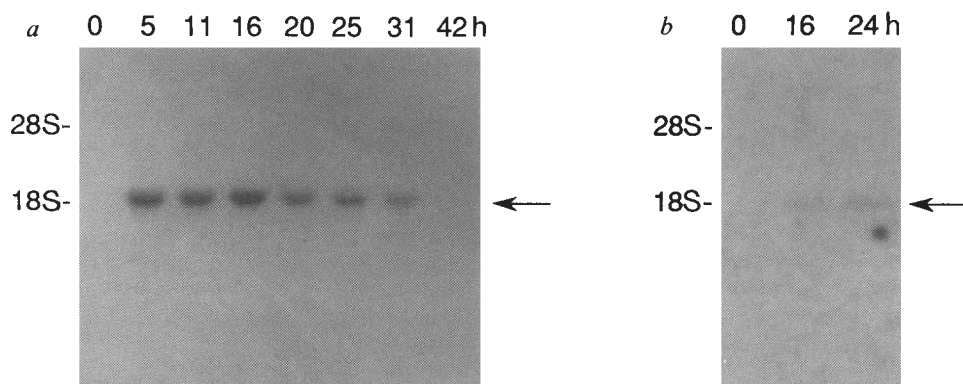
IL-1-induced PGE₂ synthesis in fibroblasts is inhibited by an *E. coli* lysate containing human IL-1ra. Human foreskin fibroblast (FF) cells (obtained from Clonetics Corp.) were seeded at 1×10^5 cells per well of a 96-well microtitre plate in RPMI-1640 medium (Mediatech) + 10% fetal bovine serum (FF medium). After 16–20 h, the cells were washed and 100 μ l lysate or lysate diluted in FF medium was added. Lysate dilutions are indicated by the asterisks. IL-1 β (100 μ l; Genzyme) at 1 U ml⁻¹ in FF medium were then added to each well. After 6 h of incubation at 37 °C, the culture supernatant was removed and assayed for PGE₂ by a competitive enzyme-linked immunosorbent assay (ELISA) as described³². PGE₂ values ($\pm$ s.d.) are given in ng ml⁻¹. Estimates of error are based on at least two determinations. *E. coli* strains and methods for preparing lysates are described in Fig. 4b. IL-1 β alone induces >63 ng ml⁻¹ PGE₂. Untreated cells produce <0.3 ng ml⁻¹ PGE₂.

sequences that are close to the consensus sequence for translation initiation¹⁴. The N-terminal 25 amino acids closely resemble a classical signal sequence, with a charged residue very near the N terminus, followed by a long stretch of hydrophobic amino acids and ending with an amino acid having a small side chain¹⁵. Cleavage by leader peptidase between the 25th and 26th amino acid encoded by the ORF would result in a mature protein of 152 amino acids (M_r 17,115) with a predicted isoelectric point of 5.2, and an N-terminal sequence corresponding to that of the monocyte-derived receptor antagonist¹². The IL-1ra protein sequence contains one N-linked glycosylation site, in accordance with the finding that two of the three forms of the antagonist produced by monocytes are glycosylated. The sequence shows limited homology to those of both IL-1 alpha (α) and IL-1 beta (β) (see below).

Evidence that the clone codes for the IL-1ra

The above data show a close correspondence between the predicted properties of the protein encoded by the cDNA and those of the monocyte-derived IL-1ra. To confirm that the cDNA codes for the antagonist, we expressed the cDNA in *Escherichia coli* and characterized the M_r and activity of the recombinant protein. The vector that we used to express the protein is derived from the phage T7 promoter-polymerase expression system

FIG. 3 Northern-blot analysis of RNA from induced monocytes and U937 cells. *a*, Monocytes were induced for various times, as indicated at the top of each lane, by plating on IgG-coated culture dishes as previously described^{11,12}. RNA was prepared as described³⁴. Each sample of total RNA (20 μ g) was electrophoresed on a 1.2% formaldehyde-agarose gel⁴². The gel was stained with ethidium bromide, allowing the 18S and 28S rRNA bands to be seen. The RNA was blotted onto nitrocellulose and probed with 2 pmol [32 P]-labelled oligonucleotide, 5'-GATTCTGAAGGCTTGCATCTT-3', at 55 °C in 10 ml 6 $\times$ NET (1 $\times$ NET is 0.15 M NaCl, 1 mM EDTA, 15 mM Tris HCl, pH 8.3) 0.1% SDS, 5 $\times$ Denhardt's solution, 100 μ g ml⁻¹ *E. coli* transfer RNA³⁸. Overnight hybridization was followed by four washes at 30 °C and two washes at 55 °C in 6 $\times$ SSC, 0.1% SDS. The blot was then exposed overnight to X-ray film. The positions of the 18S and 28S RNA bands on the blot are indicated. The arrow indicates



the IL-1ra mRNA band. *b*, U937 cells obtained from ATCC were induced with PMA (10 μ g ml⁻¹) and incubated for various times as indicated at the top of each lane of the gel. RNA was prepared, electrophoresed through agarose, and analysed by northern blotting as described in *a*, except that 10 μ g total RNA were loaded on the gel, and the blot was exposed to X-ray film for 120 h.

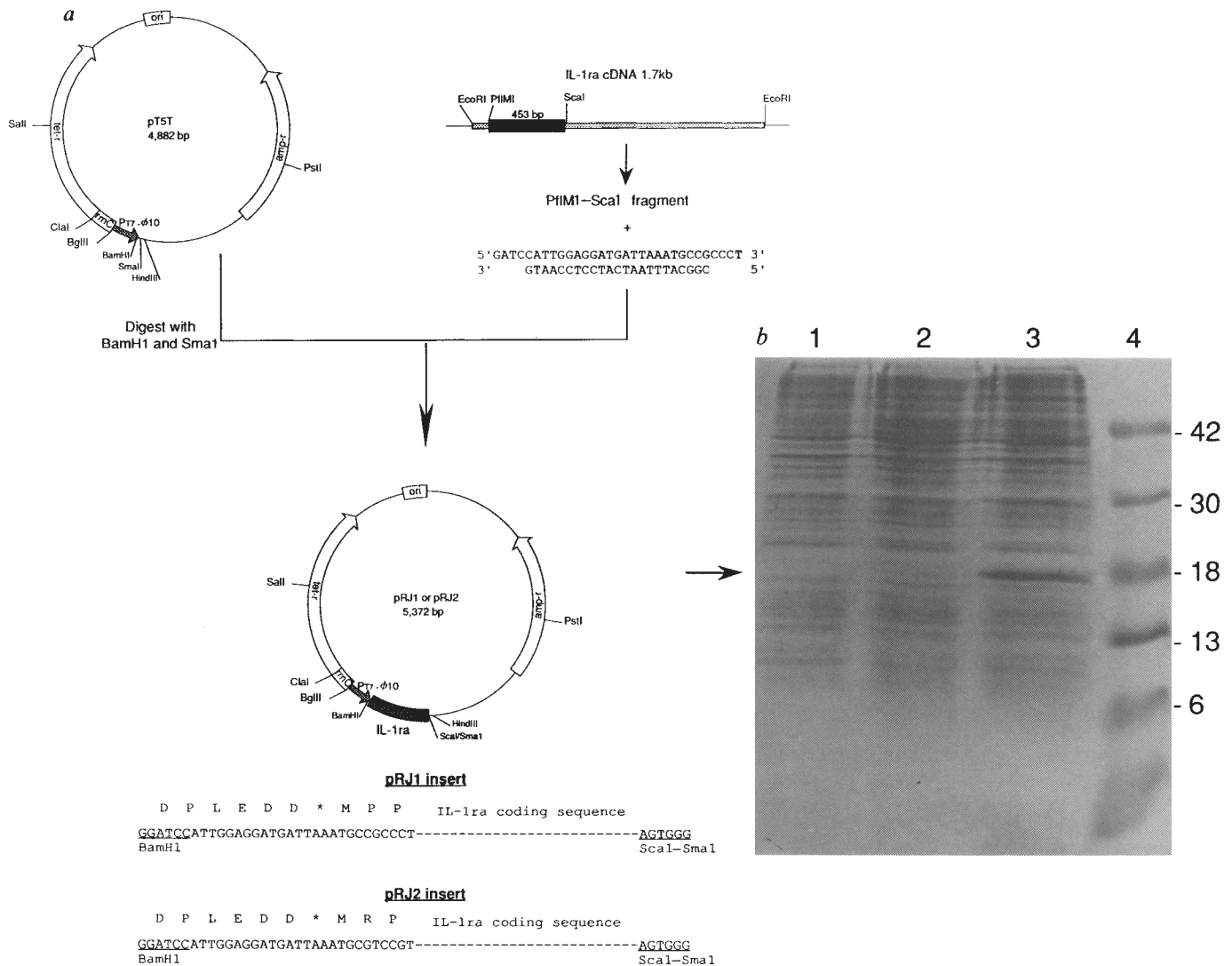


FIG. 4 Expression of recombinant IL-1ra in *E. coli*. **a**, The IL-1ra cDNA was inserted into the T7 expression vector pT5T, which is identical to plasmid pJU1003¹⁸ except for the insertion of the *E. coli* *rrnC* transcriptional terminator (*rrnC*, ref. 43) contained on the synthetic oligonucleotides

5' ATGATA AGCTGTCAAA CATGAGAATT GAGCTCCCCG GAGATCCTTA
3' GCTACTAT TCGACAGTTT GTACTCTTAA CTCGAGGGGC CTCTAGGAAT

GCGAAAGCTA AGGATTTTTT TTA 3'
CGCTTTCGAT TCCTAAAAAA AATCTAG 5'

between the unique *Bgl*II site and the *Clal* site at the 5' end of the tetracycline resistance (*tet-r*) gene. A *Pst*I-ScaI fragment derived from the IL-1ra cDNA (encoding amino acids 4-152) was inserted into *Bam*HI-SmaI-digested pT5T using a *Bam*HI-PstI oligonucleotide adaptor (shown) containing sequences for translational coupling and the first three amino acids. The derived plasmid, pRJ1, encodes a form of IL-1ra with an N-terminal sequence Met-Pro-Pro-Ser- rather than Arg-Pro-Ser-, which is the N-terminal sequence of the natural protein (see Fig. 2c, and ref. 12). The IL-1ra gene in plasmid pRJ1 was mutated to change the coding from Met-Pro-Pro-Ser- to Met-Arg-Pro-Ser- by cloning the *Bam*HI-PstI IL-1ra fragment of pRJ1

into M13mp19¹³ and performing oligonucleotide site-directed mutagenesis as described⁴⁴. The sequence of the mutagenic oligonucleotide was 5'-TGATTAATGCGTCCGTCTGGAG-3'. After mutagenesis, the *Bam*HI-PstI fragment was ligated into plasmid pT5T to yield the expression plasmid pRJ2. Abbreviations: amp-r, ampicillin resistance; ori, origin of plasmid replication; Prr- ϕ 10, T7 promoter and start of the ϕ 10 gene¹⁷; bp, base pair. **b**, Expression of recombinant IL-1ra in *E. coli*. Either plasmid pRJ2 or plasmid pT5T was used to transform *E. coli* strain BL21 (DE3)¹⁶, which contains the phage T7 RNA polymerase gene under control of the inducible *lac* operon promoter. Expression of recombinant IL-1ra was achieved by growing the strain BL21 (DE3)-pRJ2 to mid-logarithmic phase, adding isopropyl β -D-thiogalactopyranoside (IPTG) to 1 mM, and incubating for 4 h. The control strain, BL21(DE3)-pT5T, was treated similarly. To prepare lysates, cells were collected by centrifugation, suspended in $\frac{1}{10}$ vol. of 50 mM Tris-HCl buffer, pH 7.0, and passaged twice through a french pressure cell at 20,000 lb in². Protein (5 μ g) from each lysate were electrophoresed on an 18% SDS-polyacrylamide gel³³, which was then stained with Coomassie brilliant blue. Lane 1, BL21(DE3)-pT5T with IPTG; lane 2, BL21(DE3)-pRJ2 without IPTG; lane 3, BL21(DE3)-pRJ2 with IPTG; lane 4, *M*_r markers. Arrow indicates the IL-1ra band. Abbreviation: bp, base pair.

originally developed by Studier and co-workers^{16,17}, and modified by Squires *et al.*¹⁸. The system links the expression of the mature antagonist (without the signal sequence) to that of the phage T7 ϕ 10 gene by a translational coupler (Fig. 4a). Figure 4b is a photograph of a Coomassie blue-stained gel to which we applied either control *E. coli* extract or extracts from *E. coli* in which the cDNA is expressed. There is a protein

band present in the latter extracts and not in control extracts that has a mobility equal to that of the nonglycosylated IL-1 antagonist (form x) produced by monocytes¹². We further analysed bacterial extracts for their ability to prevent IL-1-induced synthesis of prostaglandin E₂ (PGE₂) by foreskin fibroblasts (Table 1). Extracts from bacteria expressing the cDNA clearly inhibited the fibroblast response to IL-1 with a similar activity

to that of the natural IL-1ra¹², whereas control extracts had no inhibitory activity. This result shows that the IL-1ra cDNA clone codes for the IL-1 receptor antagonist.

In addition, the complete IL-1ra cDNA has been inserted into a mammalian cell expression vector downstream of the human cytomegalovirus early promoter (J. Vannice *et al.*, manuscript in preparation). After COS cell transfection, an IL-1 receptor antagonist activity is detected in the medium, indicating that the cDNA codes for the antagonist pre-protein, with a functional signal sequence.

IL-1ra mRNA induction in human cells

By northern-blot analysis, we found that the induction of IL-1ra mRNA in monocytes is rapid. It is detectable 1–2 h after plating cells on IgG (data not shown) and reaches a peak concentration 4–5 h after induction (Fig. 3a). This level is maintained for about 12 h and then gradually decreases. We predicted this time course from the SDS-PAGE analysis of the [³⁵S]methionine pulse-labelling experiment (see above). IL-1ra mRNA is also produced by the U937 cell line derived from transformed human monocytes. The mRNA is induced by phorbol myristate acetate (PMA; Fig. 3b), although the rate of induction and maximal level of this message both seem to be lower than observed in human IgG induced monocytes. IgG, lipopolysaccharide and zymosan A can also induce this message in U937 cells, although not as efficiently as PMA (data not shown).

Structural relationship to IL-1

How does IL-1ra inhibit the activity of IL-1? Hannum *et al.*¹² show that radiolabelled recombinant IL-1ra binds to the IL-1 receptor on EL-4 6.1 cells with an affinity close to that of IL-1 α and IL-1 β . It therefore seems reasonable to expect that IL-1

and IL-1ra have similar structures, at least in the region(s) recognized by the receptor. We compared the sequence of IL-1ra with that of human IL-1 α and human IL-1 β ¹⁹ and found 26% homology between the antagonist and IL-1 β , and 19% homology between the antagonist and IL-1 α (Fig. 5a). Identical or similar residues between the three sequences are candidates for amino acids that are important for receptor binding. Those residues in common between IL-1 α and IL-1 β , but different in the IL-1ra sequence could be important for agonist activity.

Discussion

Control of IL-1 action could occur through regulation of its synthesis at the level of transcription, translation, protein processing or secretion, as previously proposed^{20–23}. But, there are circumstances in which it could be advantageous to control IL-1 effects post-synthetically by inhibiting its ability to interact with cellular targets. The isolation of a monocyte-derived IL-1 inhibitor that is a receptor antagonist¹² is experimental support for this mechanism. The physiological relevance of this mechanism of IL-1 regulation is supported by the finding that the antagonist is made by the same cells that make IL-1⁹, that its synthesis is induced *in vitro* by immune complexes which are a known pathological signal, and that levels of an inhibitor of IL-1 seem to be elevated in the urine of patients with pathologies including fever and myelomonocytic leukaemia (ref. 24, see below).

Cloning and expression of a cDNA for IL-1ra should aid the determination of the role of IL-1 in diseases because only disease processes involving IL-1 should be modified by this specific¹¹ antagonist. In addition, attempts to study IL-1-mediated diseases by correlating IL-1 levels with the disease state will be helped by simultaneous determinations of the levels of IL-1 and

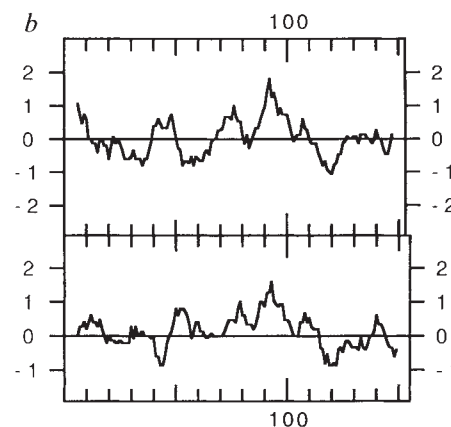
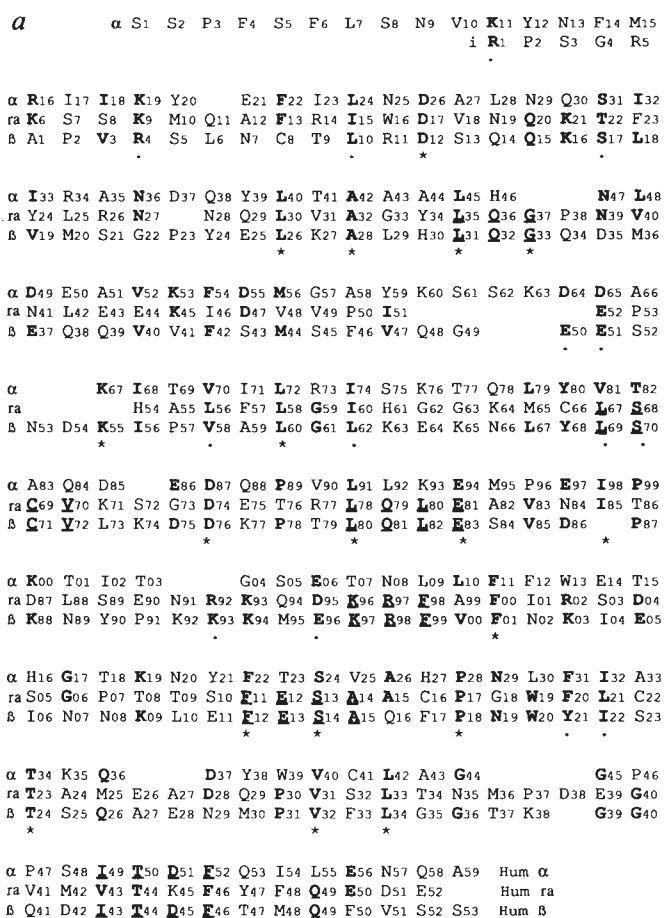


FIG. 5 Structural similarity between IL-1 and IL-1ra. **a**, Alignment of human IL-1 α , IL-1 β and IL-1ra. Alignment between the mature sequences of human IL-1 α and IL-1 β is based on the method of Young and Sylvester²⁵. IL-1ra is aligned to IL-1 β according to the Beckmann Microgenie program using the normal protein homology parameters, except that gaps in the alignment are limited to turns between the beta strands of the IL-1 β structure²⁶. Homologous or similar residues between at least two sequences are in bold. Homologous residues in all sequences are indicated by an asterisk. Similar residues in all sequences are indicated by a dot. Short stretches of homology (at least three consecutive homologous residues) between either IL-1 β and IL-1ra, or IL-1 α and IL-1 β , are underlined. Hum, human. **b**, Comparison of hydrophilicity plots for IL-1 β and IL-1ra. Hopp and Woods hydrophilicity plots⁴⁵ were drawn using the DNA Strider program written for the Macintosh computer⁴⁶. Positive values on the y-axis indicate hydrophilic regions. The x-axis refers to amino-acid residue number in the mature protein sequence. Upper panel, IL-1ra; lower panel, IL-1 β .

its antagonist. Such attempts are otherwise fraught with difficulty because high levels of immunologically detectable IL-1 could be biologically ineffective owing to the presence of the antagonist and, conversely, low levels of IL-1 could be effective in the absence of the antagonist. These problems can now be corrected because it is possible to prepare reagents to make simultaneous measurements of agonist and antagonist levels in diseased tissue.

A global search of all sequences in the Protein Information Resource data base using the Intelligenetics FASTDB program showed that IL-1ra and IL-1 β are 30% homologous. The Young and Sylvester²⁵ alignment (Fig. 5a) finds slightly less homology (26%) because gaps are excluded from the beta strands of the three-dimensional (3-D) crystal structure of IL-1 β ²⁶. Many of the identical residues between IL-1ra and IL-1 β occur in short stretches of amino acids and are located in the C-terminal three-fifths of the two proteins. Hydrophilicity plots for the two proteins are remarkably similar in this same region (Fig. 5b). In this region, the two molecules could have similar 3-D structures and contain the residues that are necessary for receptor binding. In fact, when the homologous residues between IL-1ra and IL-1 β are mapped onto the 3-D structure of IL-1 β ²⁶, many of them cluster around one vertex of the pyramidal molecule. Other workers have attempted to align the sequences of IL-1 α and IL-1 β to identify homologous residues on the assumption that they would be involved in receptor binding. But many of these homologies are hydrophobic residues that could be important for maintaining the native structure of the protein². The addition of a third sequence, that of IL-1ra, to these sequence comparisons should help identify important residues.

The relationship between IL-1 and IL-1ra is reinforced by the fact that the organization of the IL-1ra mRNA is similar to that of both IL-1 α and IL-1 β ¹⁹. A large ORF encoding the protein is preceded by a relatively short 5' UTR and followed by a long 3' UTR.

Although there are sufficient similarities between IL-1ra and IL-1 β to support the assertion that the antagonist is a member of the IL-1 family of molecules, there are clearly significant differences between IL-1ra and either IL-1 α or β . First, IL-1ra is an antagonist of IL-1, biological activity and as such, cannot proceed through one or more of the steps in IL-1 signal transduction after binding to the receptor: these could include internalization and nuclear association²⁹, and activation of transcription factors leading to the induction of certain genes³⁰. The structural

elements leading to this difference are not yet known but comparison of the sequences of these molecules could provide a clue. Discovering this difference will be important because natural proteins that block the action of other proteins by competitively binding to their receptor without initiating a normal response have not been previously reported.

Second, IL-1ra differs from IL-1 α or IL-1 β in its mechanism of secretion. From the sequence or the cDNA it is predicted that IL-1ra is synthesized as a pre-protein consisting of a classical signal sequence fused to the mature active protein. IL-1 α and IL-1 β are also secreted proteins but they are synthesized as much larger precursors without obvious signal sequences, from which either 112 or 116 amino-acid residues are removed during processing^{2,23}. IL-1ra therefore apparently uses the classical pathway involving a signal sequence, signal recognition particle and leader peptidase, whereas IL-1 α and IL-1 β seem to be transported and processed by other so far unidentified mechanisms.

Third, the 3' UTR of the IL-1ra mRNA, although it is extremely A-U rich, lacks the sequence AUUUA. This pentanucleotide sequence has been implicated in regulating the half-life of mRNAs of many proteins that control the growth of cells, such as IL-1 α , IL-1 β , granulocyte macrophage colony-stimulating factor and tumour necrosis factor- α ²².

One other IL-1 inhibitor has been reported that has also been characterized as an IL-1 receptor antagonist. This is the urine-derived IL-1 inhibitor that has been partially purified and characterized by Seckinger *et al.*^{10,31}. There are good reasons to believe that this protein is closely related to the monocyte-derived receptor antagonist. The M_r s, isoelectric points and modes of action of the two proteins are similar, and we have recently identified the monocyte-derived antagonist in urine by western blotting with a polyclonal antibody against purified recombinant IL-1ra (data not shown).

We have cloned the cDNA for the human IL-1 receptor antagonist, IL-1ra, and expressed the protein in *E. coli*. Further study of this protein and its tissue location should be valuable in determining the role of IL-1 in human physiology. The protein should also be useful for studying the interactions between IL-1 and its receptor, and could also help elucidate the mechanism by which IL-1 triggers many cellular responses. Finally, the IL-1 antagonist is potentially useful as a therapeutic agent for the treatment of IL-1-mediated diseases in humans, in particular, inflammatory disorders such as rheumatoid arthritis. □

Received 1 December; accepted 19 December 1989.

- Dinarello, C. A. *FASEB J.* **2**, 108-115 (1988).
- Maliszewski, C. R. *et al. Molec. Immun.* **25**, 429-437 (1988).
- Oppenheim, J. J., Kovacs, E. J., Matsushima, K. & Durum, S. K. *Immun. Today* **7**, 45-56 (1986).
- Eastgate, J. A. *et al. Lancet* **ii**, 706-709 (1988).
- Pettipher, E. R., Higgs, G. A. & Henderson, B. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8749-8753 (1986).
- Okusawa, S., Gelfand, J. A., Ikejima, T., Connolly, R. J. & Dinarello, C. A. *J. clin. Invest.* **81**, 1162-1172 (1988).
- Dayer, J.-M. and Seckinger, P. in *Interleukin 1—Inflammation and Diseases* (eds Borford, R. H. R. & Henderson, B.) (Elsevier Science Publishers) (in the press).
- Larick, J. W. *Immun. Today* **10**, 61-66 (1989).
- Arend, W. P., Joslin, F. G. & Massoni, R. J. *J. Immun.* **134**, 3868-3875 (1985).
- Seckinger, P., Lowenthal, J. W., Williamson, K., Dayer, J.-M. & MacDonald, H. R. *J. Immun.* **139**, 1546-1549 (1987).
- Arend, W. P., Joslin, F., Thompson, R. C. & Hannum, C. H. *J. Immun.* **143**, 1851-1858 (1989).
- Hannum, C. H. *et al. Nature* **343**, 336-340.
- Yanisch-Perron, C., Vieira, J. & Messing, R. *Gene* **33**, 103-119 (1985).
- Kozak, M. *Cell* **44**, 283-292 (1986).
- von Heijne, G. *Nucleic Acids Res.* **14**, 4683-4690 (1986).
- Studier, F. W. & Moffat, B. *J. molec. Biol.* **189**, 113-130 (1986).
- Rosenberg, A. *et al. Gene* **56**, 125-135 (1987).
- Squires, C. Y., Childs, J., Eisenberg, S. P., Polverini, P. J. & Sommer, A. *J. biol. Chem.* **263**, 16297-16302 (1988).
- March, C. J. *et al. Nature* **315**, 641-647 (1985).
- Fenton, M. J. in *Hematopoiesis UCLA Symposia Series Vol. 120* (eds Clark, S. & Golde, D.) (Alan R. Liss, New York) (in the press).
- Arend *et al. J. Immun.* **143**, 118-126 (1989).
- Malter, J. S. *Science* **246**, 664-666 (1989).
- Black, R. A., Kronheim, S. R. & Sleath, P. R. *FEBS Lett.* **247**, 386-390 (1989).
- Balavoine, J.-F. *et al. J. clin. Invest.* **78**, 1120-1124 (1986).
- Young, P. R. & Sylvester, D. *Protein Engng.* **2**, 545-551 (1989).
- Priestle, J. P., Schaefer, H. P. & Grütter, M. G. *EMBO J.* **7**, 339-343 (1988).
- Dower, S. K. *et al. Nature* **324**, 266-268 (1986).
- Kilian, P. L. *et al. J. Immun.* **136**, 4509-4514 (1986).
- Mizel, S. B., Kilian, P. L., Lewis, J. C., Paganelli, K. A. & Chizzonite, R. A. *J. Immun.* **138**, 2906-2912 (1987).
- Muegge, K. *et al. Science* **246**, 249-251 (1989).
- Seckinger, P. *et al. J. Immun.* **139**, 1541-1545 (1987).
- Pradelles, P., Grassi, J. & Maclouf, J. *Analyt. Chem.* **57**, 1170-1173 (1985).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156-159 (1987).
- Aviv, H. & Leder, P. *Proc. natn. Acad. Sci. U.S.A.* **69**, 1408-1412 (1972).
- Huynh, T. V., Young, R. A. & Davis, R. W. in *DNA Cloning Vol 1* (ed. Glover, D. M.) 49-78 (IRL, Washington DC, 1985).
- Gübler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Wallace, R. B. & Miyada, C. G. *Meth. Enzym.* **152**, 432-442 (1987).
- Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767-4771 (1987).
- Lehrach, H., Diamond, D., Wozney, J. M. & Boedtker, H. *Biochemistry* **16**, 4743-4747 (1977).
- Normanly, J., Ogden, R. C., Horvath, S. J. & Abelson, J. *Nature* **321**, 213-219 (1986).
- Kunkel, T. A., Roberts, J. D. & Zakour, R. A. *Meth. Enzym.* **154**, 367-382 (1987).
- Hopp, T. P. & Woods, K. R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3824-3828 (1981).
- Marck, C. *Nucleic Acids Res.* **16**, 1829-1836 (1988).

ACKNOWLEDGEMENTS. We thank David Hirsh and Anne Burkholder for critically reviewing the manuscript, Patricia Heimdal, Carol Wilcox, Bob Weaver and Fenneke Joslin for technical assistance, and Tim Klein for helping with figures.

Dark matter and the age of globular clusters

David Dearborn*, Georg Raffelt†, Pierre Salati†‡, Joseph Silk† & Alain Bouquet§

* Lawrence Livermore National Laboratory, Livermore, California 94550, USA

† Astronomy Department, Center for Particle Astrophysics, and

‡ Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, USA

§ Laboratoire d'Annecy le Vieux de Physique des Particules, 74941 Annecy le Vieux Cedex, France

GALACTIC dark matter may consist of weakly interacting particles which can be captured and trapped in stars¹, and which would then contribute to the transfer of energy^{2,3}. A special class of these particles ('cosmions'), with weak cross-sections that are larger than standard has been invoked as a solution of the solar-neutrino problem, and also as a means of suppressing convection in the cores of horizontal-branch stars⁴. Here we investigate this latter effect numerically and find that the convection breaking in horizontal-branch stars by cosmions, or by any other novel mode of energy transfer, induces thermal relaxation oscillations with a period of $\sim 5 \times 10^5$ yr, corresponding to the core Kelvin-Helmholtz time-scale. These thermal pulses can be understood analytically in terms of a simple two-zone model. Observationally, the brightness and brightness dispersion of horizontal-branch stars increases, the periods of RR Lyrae stars change over a pulsation timescale, and the duration of central helium burning slightly decreases. None of these effects is in conflict with observations, but, on the contrary, point to a speculative resolution of the age problem for globular clusters and to an alternative explanation of the period fluctuations of RR Lyrae stars.

In the standard theory of stellar evolution, horizontal-branch (HB) stars are thought to be configurations with a dense core ($\rho \approx 10^4$ g cm⁻³) and a large envelope. At the interface between core and envelope, hydrogen burns in a thin shell, whereas in the centre of the core, helium burning provides a second energy source. When helium is exhausted in the centre, the star enters a double-shell configuration and steadily increases its luminosity: it ascends the asymptotic giant branch (AGB). In Fig. 1, the dotted line is the absolute bolometric magnitude, as a function of time, for a typical HB star with mass $M = 0.8 M_\odot$ (solar masses), metallicity $Z = 0.004$ and initial core mass of $0.46 M_\odot$. The inner core is convective, initially $M_{\text{conv}} \sim 0.12 M_\odot$, dredging more helium fuel into the localized nuclear burning region than would be available otherwise. Renzini⁴ predicted that convection breaking by cosmions would lead, therefore, to an observable reduction of the HB lifetime.

To test this prediction we have numerically simulated the evolution of an HB star, taking cosmion energy transfer into account. We assume that cosmions are steadily accreted during the preceding main-sequence phase, that they do not substantially change the main-sequence and red-giant evolution, and that they are not ejected during the helium flash—the violent

expansion of the core associated with helium ignition. For sufficiently large cosmion interaction cross-sections, the energy transfer is conductive ($\propto N_x \sigma_x^{-1}$, where N_x is the total number of cosmions in the star and σ_x is the cosmion interaction cross-section), whereas for small cross-sections it is non-local ($\propto N_x \sigma_x$). Initially, when the core consists only of helium, the turnover between the two regimes is at $\sigma_{\text{He}} \sim 0.36 \times 10^{-36}$ cm² $A_x^{1/2}$ where A_x is the cosmion 'atomic weight' and σ_{He} is the scattering cross-section on helium. If cosmions solve the solar-neutrino problem⁵, the cross-section must be so large that the conduction approximation is rigorously justified for HB stars. We use recently calculated⁶ thermal conduction coefficients and assume coherent interactions where the scattering cross-section on a nucleus with atomic weight A_n is $\sigma_n \propto A_x^2 A_n^4 (A_x + A_n)^{-2}$ (ref. 7), assume a cosmion mass A_x of 4, and assume the parameter that governs the efficiency of energy transfer, N_x/σ_{He} , to be 3×10^{81} cm⁻².

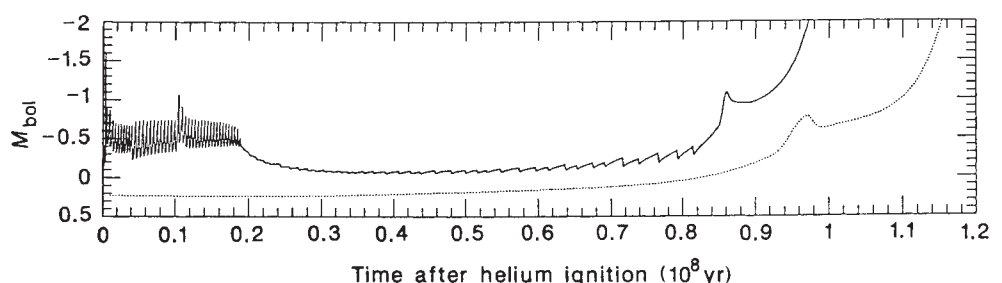
Surprisingly, there exists no stable equilibrium solution of the stellar structure equations: the star exhibits regular thermal pulses until cosmions are no longer able to break convection, a saturation effect that occurs because of our assumed coherent interaction law, which entails a reduction in the efficiency of energy transport as carbon builds up. The luminosity evolution is shown in Fig. 1 as a solid line. The behaviour of the star during these pulses is best described using Fig. 2 where the stellar energy sources, central density, central temperature and convective core mass are plotted for one period of oscillation. The stellar core slowly contracts, and the central density and temperature rise, until central helium burning sharply rises. The energy can no longer be carried by cosmions and convection commences. Simultaneously, the core rapidly expands: most of the helium energy production is absorbed by this process and the surface luminosity increases only mildly. This increase is further moderated by the momentary decrease of the hydrogen shell source caused by the core expansion. After this expansion, helium burning is quenched, convection ceases, and the core begins to contract again, starting a new cycle. The short bursts of convection extend to larger radii than does steady convection in a standard HB star and so the amount of helium available for core burning exceeds that in the standard case.

The thermal pulses can be understood as thermal relaxation oscillations if one adopts a simple model in the spirit of Buchler and Perdang's⁸ explanation of thermal pulses of AGB stars. We divide the core of the HB star into two zones. The first is the central nuclear burning region with mass $M_1 \approx 0.05 M_\odot$, temperature $T_1 \approx 10^8$ K and a positive heat capacity c_1 . The nuclear energy generation rate ϵ_{He} varies $\sim T_1^{40}$. The second zone is the remainder of the core with mass $M_2 \approx 0.5 M_\odot$ and constitutes a gravothermal buffer with negative heat capacity $-c_2$. The temperature evolution in the two regions is modelled by a set of nonlinear differential equations,

$$\begin{aligned} \dot{T}_1 &= (+c_1 M_1)^{-1} [\epsilon_{\text{He}}(T_1) M_1 - K_1 (T_1^4 - T_2^4)] \\ \dot{T}_2 &= (-c_2 M_2)^{-1} [K_1 (T_1^4 - T_2^4) - K_2 (T_2^4 - T_{\text{ex}}^4)] \end{aligned} \quad (1)$$

where T_{ex} is the temperature outside of the core, K_1 accounts for the energy transfer between zones 1 and 2 by means of

FIG. 1 Brightness evolution of a standard HB star, $M = 0.8 M_\odot$, $Z = 0.004$ (dotted line), and the same star with cosmion energy transfer, $A_x = 4$, $N_x/\sigma_{\text{He}} = 3 \times 10^{81}$ cm⁻² (solid line). Note that $M_{\text{bol}} = 4.72 - 2.5 \log_{10} (L_{\text{surface}}/L_\odot)$.



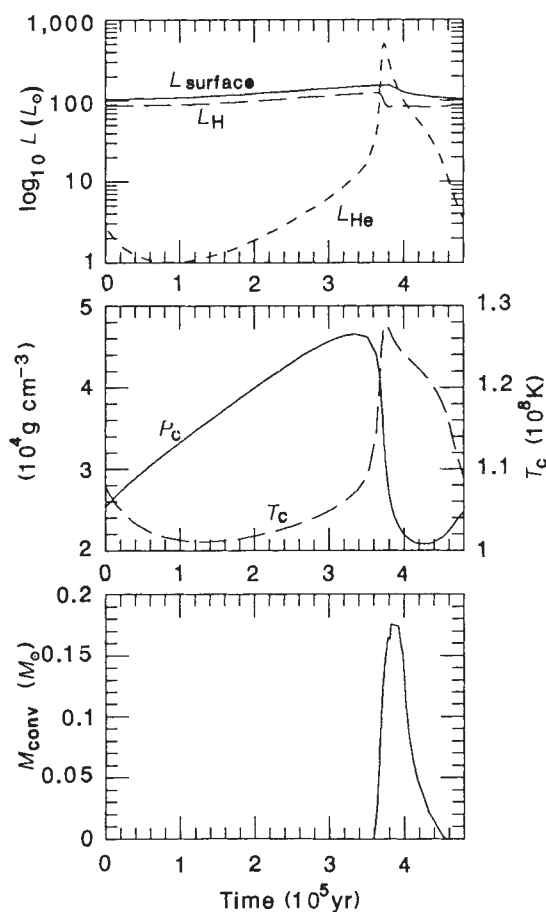


FIG. 2 Helium burning luminosity L_{He} , luminosity of the hydrogen shell source L_{H} , surface luminosity L_{surface} , central density ρ_c , central temperature T_c and size of the convective core M_{conv} for one cycle of thermal pulsing.

radiative and cosmion transfer, and K_2 is the thermal conductance between the core and the envelope. This set of equations is supplemented by the condition that when $T_1 - T_2$ exceeds a critical value $\nabla_{\text{ad}} T$, convection transports energy so efficiently that the temperature difference remains fixed at this value and the nuclear burning energy is directly dumped into the gravothermal buffer. This simple model is sufficient to account for our numerical results: if K_1 is chosen too small to break convection, equation (1) has a fixed point $(T_1, T_2)_{\text{fix}}$ corresponding to the usual convective equilibrium solution of stellar structure. If K_1 is so large that convection is broken, equation (1) has a limit cycle, $[T_1(t), T_2(t)]_{\text{limit}}$, which can be parameterized as a function of time, with no stable fixed point.

To discuss the observational consequences of our result, we have plotted (Fig. 3) the HB luminosity function, that is, the number of stars per interval of brightness, for the standard model (dotted histogram) and for the stellar model with cosmion energy transfer (solid histogram). The luminosity dispersion of the HB in globular clusters is observed to be between 0.2 and 0.7 mag (A. Sandage, manuscript in preparation), and does not allow one to distinguish between our two cases. The reduction of the HB lifetime of $\sim 20\%$ seems to be too small to be in serious conflict with number counts in globular clusters. Our pulses would also lead to a periodic change in the oscillation periods of RR Lyrae stars. For many of these stars, the periods change at a rate $\dot{P} \approx 10^{-9}$, corresponding well to what is expected on the basis of our calculations. For a conventional explanation involving random mixing events in the semi-convective layer and for references to the observational data see ref. 9. We have

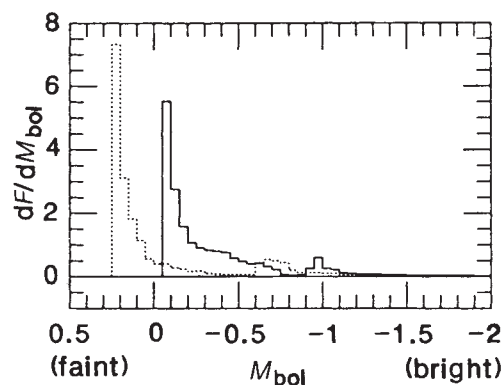


FIG. 3 Luminosity function of the HB for our standard model (dotted histogram) and our stellar model with cosmions (solid histogram). The HB luminosity of our standard stellar model is somewhat high because we have ignored mass-loss during the red-giant evolution, leading to a somewhat large mass ($0.8 M_{\odot}$) on the HB.

not investigated the effect of mass-loss during the red-giant phase on the distribution of pulsing stars along the HB. Removing enough mass from the envelope to shift the surface temperature to the RR Lyrae instability strip, however, led to large temperature oscillations.

An increase of the true HB luminosity would imply that globular clusters are farther away than had been thought previously, but a brightness increase by ~ 0.3 mag as in our example corresponds to a distance increase of $\sim 15\%$, within the uncertainty of other distance determinations. The revised distances require, however, a revision of the absolute brightness of the main-sequence turnoff points and hence of the globular cluster ages. Leaving all other parameters unchanged, a shift of the HB absolute magnitude leads to a change of the inferred cluster age by $\delta \log_{10}(\text{age}/10^9 \text{ yr}) = 0.41 \delta M_{\text{bol}}$ (ref. 10). If cosmions contributed to the energy transfer at the level assumed in our calculation, $\delta M_{\text{bol}} = 0.3$ mag and the cluster ages would have been overestimated by $\sim 30\%$. Globular cluster ages are notoriously difficult to reconcile with the standard cosmological parameters of the Friedmann universe unless extreme values are adopted, such as $q_0 \leq 0.1$ and $H_0 \leq 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Also, the cosmion reduction for the inferred ages of the youngest globular clusters brings these values closer to the local galactic disk age of $\sim 9 \times 10^9$ yr, inferred from the white-dwarf luminosity function¹¹.

Cosmion accretion rates should vary with stellar location: for example, stars in globular clusters in the inner galaxy, stars in the cores of globular clusters (if the dark matter is locally confined) or nearby dwarf spheroidals should accrete substantially more cosmions than stars in globulars at large galactocentric radii or stars in the outer regions of globulars. This may eventually lead to a unique observational signature of this model. $\square$

Received 14 August; accepted 22 November 1989.

1. Press, W. H. & Spiegel, D. N. *Astrophys. J.* **296**, 679–684 (1985).
2. Steigman, G., Sarazin, C. L., Quintana, H. & Faulkner, J. *Astr. J.* **83**, 1050–1061 (1978).
3. Spiegel, D. N. & Press, W. H. *Astrophys. J.* **294**, 663–673 (1985).
4. Renzini, A. *Astr. Astrophys.* **171**, 121–122 (1987).
5. Gilliland, R. L., Faulkner, J., Press, W. H. & Spigel, D. N. *Astrophys. J.* **306**, 703–709 (1986).
6. Gould, A. & Raffelt, G. *Astrophys. J.* (in the press).
7. Goodman, M. W. & Witten, E. *Phys. Rev. D* **31**, 3059–3063 (1985).
8. Buchler, J. R. & Perdang, J. *Astrophys. J.* **231**, 524–533 (1979).
9. Sweigart, A. V. & Renzini, A. *Astr. Astrophys.* **71**, 66–78 (1979).
10. Iben, I. Jr & Renzini, A. *Phys. Rep.* **105**, 329–406 (1984).
11. Winget, D. E. et al. *Astrophys. J.* **315**, L77–L81 (1987).

ACKNOWLEDGEMENTS. We thank E. Belserene, J. Buchler, P. Eggleton, I. Iben, I. King, R. Romani, V. Trimble and R. Ulrich for discussions and comments. G.R. acknowledges an Otto-Hahn fellowship of the Max-Planck-Society, P.S. a fellowship of the Miller Institute for Basic Research in Science. In Berkeley, this work was supported, in part, by grants from NASA, the Department of Energy, the NSF and the Institute for Geophysics and Planetary Physics (Livermore).

PSR1737–30 and period discontinuities in young pulsars

J. McKenna & A. G. Lyne

University of Manchester, Nuffield Radio Astronomy Laboratories,
Jodrell Bank, Macclesfield, Cheshire SK11 9DL, UK

THE rotation rate of pulsars decreases as kinetic energy is lost by the emission of electromagnetic radiation and particles. This gradual spin-down is sometimes interrupted, however, by abrupt increases in rotational frequency. Such discontinuities—known as glitches—arise because of structural changes in neutron stars, and can provide valuable information on their interiors. Glitches generally occur more frequently in younger pulsars¹. A timing program has been monitoring the periods of the forty predominantly young pulsars found in the Jodrell Bank 1,400-MHz survey^{2,3,4}. As part of this study, we report here the observation that one pulsar, PSR1737–30, has undergone no fewer than five glitches in the past three years. Although not especially young, this pulsar is glitching at a rate nearly an order of magnitude greater than any other known pulsar. We speculate that the higher temperatures associated with the youngest pulsars might prevent the build-up of the stresses the sudden release of which is the cause of glitches. If so, this might explain the lower glitch activity now evident in those pulsars that are less than 2,000 years old.

Pulse arrival times for the forty pulsars were obtained using the 76-m Lovell telescope at Jodrell Bank in conjunction with dedispersing receivers centred on frequencies ranging from 606 to 1,666 MHz. The arrival times were corrected to the Solar System barycentre using the JPL DE 200 ephemeris^{5,6}. Timing residuals were then obtained by subtracting from the observed arrival times those expected from a simple spin-down model. Improved estimates of a pulsar's rotation parameters were made by fitting to a reasonable span of these residuals⁷. In addition to PSR1737–30, the timing program has uncovered an unusual binary pulsar, PSR1820–11 (ref. 8), and revealed a glitch in each of the pulsars PSR1736–29, PSR1823–13 and PSR1859+07 (A.G.L. and J.M., manuscript in preparation).

The behaviour of the period of PSR1737–30 over a three-year interval is shown in Fig. 1. Most of the period P and its first derivative $\dot{P}$ have been removed to reveal glitches that would otherwise be swamped by the pulsar's rapid spin-down. The figure shows five discontinuities, each of them indicating a decrease in the pulse period. The four largest of these glitches are quite distinct whereas the fifth, which occurred between 4 June 1988 and 6 July 1988, is small but significant.

The timing residuals between glitches were fitted with six slightly different sets of P and $\dot{P}$. Adjacent fits were extrapolated to estimate the dates of the glitches and the fractional changes in period and period derivative ($\Delta P/P$ and $\Delta \dot{P}/\dot{P}$). These results are shown in Table 1. The large error in the quoted epoch of the first glitch arises because of possible ambiguities in pulse counting. The calculations of $\Delta P/P$ and $\Delta \dot{P}/\dot{P}$ make no allowance for any short-term features in post-glitch relaxation that might have been missed by the timing observations. The

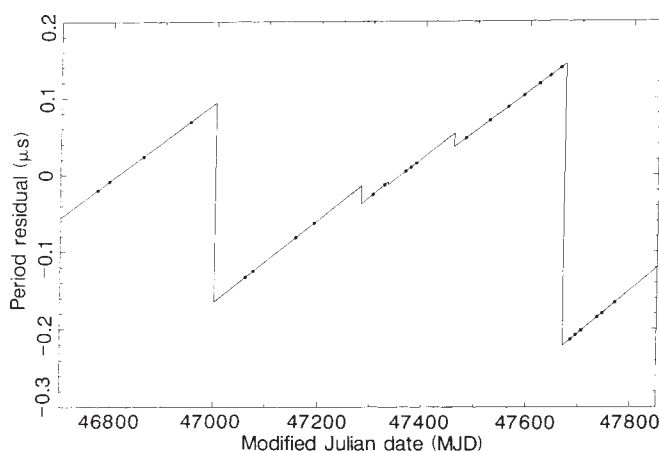


FIG. 1 A schematic representation of the period of PSR1737–30. A period of 606.59181 ms (epoch modified Julian date MJD, 47,475) and a period derivative of $460 \times 10^{-15} \text{ s s}^{-1}$ have been removed. The errors on the data points are roughly one-hundredth of the diameter of the symbols.

position used in the timing analysis, right ascension, α (1950) = $17^{\text{h}} 37^{\text{m}} 21.17 \pm 0.10^{\text{s}}$ and declination, δ (1950) = $-30^{\circ} 14' 10.0 \pm 1.5''$, was obtained at the Very Large Array (VLA)⁹ and has associated errors which could also have a significant effect on the values of $\Delta \dot{P}/\dot{P}$. These errors and the limited data coverage are reflected in the uncertainties quoted in the table. The fractional change in period displays a wide range of values and glitches that occur roughly once every six months, more frequently than in any other pulsar known. The next most frequently glitching pulsar known is the Vela pulsar, PSR0833–45, for which 8 glitches have been observed in an interval of 20 years (see, for example, ref. 10).

Glitch behaviour is generally thought to be related to the internal structure of a neutron star¹¹. It was originally proposed that glitches were the manifestation of starquakes. As a pulsar spins down, the equilibrium oblateness of its solid components decreases, causing elastic stresses to build up. When a critical strain is reached, a starquake violently reduces the oblateness and the accompanying reduction of the star's moment of inertia produces a sudden increase in the crust's angular velocity. However, if crustquakes—the cracking of the outer layers of a neutron star¹²—were the origin of the glitches in PSR1737–30, their frequency and magnitude would reduce the crust's oblateness to zero in less than 100 years, a very short time compared with the pulsar's age. Crustquakes are therefore unlikely to be the source of these glitches. Corequakes¹³, in which a fracturing solid core of the neutron star causes spin-ups, could account for the behaviour of PSR1737–30 but the assumption that neutron stars have solid cores is not widely accepted¹.

We believe that glitches are the result of the discontinuous transfer of angular momentum from the star's fluid interior to its more slowly rotating crust. The most plausible source of such behaviour is the neutron superfluid which coexists with the lattice of nuclei forming the inner layers of a neutron star's crust^{14–16}. The angular momentum of the superfluid is carried by vortices that move radially outwards as the pulsar spins down under the influence of external torques. This motion is impeded by the pinning of vortex lines to the nuclei but the pinning is not perfect and the vortices creep thermally outwards, coupling the rotation of the superfluid to that of the crust. A glitch arises when a substantial proportion of the vortices catastrophically unpin and move suddenly outwards. The angular momentum thus lost from the superfluid is transferred to the crust—the rotation of which is what we actually observe—and a sudden increase in pulse frequency is seen.

Immediately after a glitch, the outward drift of the vortices

TABLE 1 The glitches of PSR1737–30

Glitch epoch (MJD)	$\Delta P/P$ $\times 10^9$	$\Delta \dot{P}/\dot{P}$ $\times 10^3$	Period without data (MJD)
46,953–47,053	-420 ± 20	2.8 ± 0.8	46,953–47,053
47,281 ± 2	-33 ± 5	1.7 ± 4.0	47,203–47,302
47,332 ± 16	-7 ± 5	-1 ± 12	47,320–47,363
47,458 ± 2	-30 ± 8	0 ± 4	47,387–47,475
47,670.2 ± 0.2	-600.9 ± 0.6	2.0 ± 0.4	47,663–47,675

TABLE 2 The ten youngest known pulsars

PSR	Age, τ (10^3 yr)	Magnetic field (10^{12} G)	$ \dot{\Omega} $ (10^{-12} rad s $^{-2}$)	Glitch activity* (10^{-7} yr $^{-1}$)
0531+21	1.2	4	2,380	0.06
1509-58	1.5	15	429	~0
0540-69	1.7	5	1,180	~0
0833-45	11	3	98	8
1800-21	16	4	47	?
1737-30	20	17	8	4
1823-13	21	3	46	5
1930+22	40	3	17	~0
2334+61	40	10	5	~0
1727-47	80	12	1	~0

* Defined as the mean fractional change in period per year. This factor therefore depends on the size of the glitches as well as their rate of occurrence.

is much reduced and the transfer of angular momentum from the neutron superfluid to the crust essentially ceases. This produces an increase in the pulsar's period derivative which decays in an approximately exponential fashion as vortex creep resumes. A detailed study of this post-glitch decay would be a useful probe of the internal structure of a neutron star. In particular, the timescales of the relaxation could provide a means for estimating the internal temperatures of pulsars¹⁷.

So far, the limited quantity of post-glitch data for PSR1737-30 has precluded any detailed study of the relaxation but we have seen some evidence for exponential decay. The residuals after the first and fourth glitches show cubic variations too large to be due to the spin-down of an oblique rotator⁷. These could be manifestations of the third-order term in the power-series expansion of an exponential function. Unfortunately, the paucity of data means that no exponential fit to any of the pulsar's post-glitch residuals is possible.

If glitches are associated with a rapid transfer of angular momentum from the interior to the crust of a neutron star, we might expect a reduction in glitch activity as the magnitude of angular frequency derivative $\dot{\Omega}$ decreases. Some empirical evidence for this comes from statistical analyses of observed glitches¹. However, there is the suggestion even in the limited data set of Table 2 that, for the youngest pulsars, some other factor is important in determining glitch activity. In this table we formally define 'glitch activity' as the mean fractional change in period per year owing to glitches; this variable therefore depends on both the size and frequency of glitches, and provides a crude measure of the glitch behaviour. The three youngest known pulsars, PSR0531+21, PSR0540-69 (ref. 18) and PSR1509-58 (ref. 19), also have the largest known frequency derivatives but show comparatively little glitch activity. The next four pulsars are an order of magnitude older and display considerably greater activity. Although PSR1800-21 is the only one which has not shown at least one large period jump, its timing residuals strongly suggest that it has glitched in the recent past (J.M., manuscript in preparation). This clear dichotomy amongst the pulsars known to have characteristic ages $\tau \leq 20 \times 10^3$ yr, indicates that very young pulsars, $\tau \leq 2 \times 10^3$ yr, transfer angular momentum to their crusts relatively smoothly, despite their enormous frequency derivatives. A plausible explanation for this is that their higher temperatures reduce the importance of pinning and permit an easier outward drift of superfluid vortices. This could limit the development of the stresses that presumably result in glitches. Pulsars with $\tau \geq 10^4$ yr are likely to be cooler so that this effect may not be so important. Their glitch activity would then depend mainly on their frequency derivatives.

PSR1737-30 is probably the best pulsar yet discovered with which to study the behaviour of a neutron star immediately

following a glitch. Given the high frequency of glitches, this pulsar is constantly being observed in the hope that the glitch event itself may be studied. Regular timing observations will also permit a study of the post-glitch relaxation, although only for the limited time before the next glitch occurs. This object will thus provide a superb opportunity to study the densest matter observable in the universe, namely the interior of a neutron star. □

Received 23 October; accepted 19 December 1989.

1. Alpar, M. A. & Ho, C. *Mon. Not. R. astr. Soc.* **204**, 655-667 (1983).
2. Clifton, T. R. thesis, Univ. of Manchester (1985).
3. Clifton, T. R. & Lyne, A. G. *Nature* **320**, 43-45 (1986).
4. McKenna, J. in *Timing Neutron Stars* (eds Ögelman, H. & van den Heuvel, E. P. J.) 143-152 (Kluwer, Dordrecht, 1989).
5. Standish, E. M. *Astr. Astrophys.* **114**, 297-302 (1982).
6. Backer, D. C. in *Timing Neutron Stars* (eds Ögelman, H. & van den Heuvel, E. P. J.) 3-16 (Kluwer, Dordrecht, 1989).
7. Manchester, R. N. & Taylor, J. H. in *Pulsars* (Freeman, San Francisco, 1977).
8. Lyne, A. G. & McKenna, J. *Nature* **340**, 367-369 (1989).
9. Braun, R., Goss, W. M. & Lyne, A. G. *Astrophys. J.* **340**, 355-361 (1989).
10. Cordes, J. M., Downs, G. S. & Krause-Polstorff, J. *Astrophys. J.* **330**, 847-869 (1988).
11. Pines, D., Shaham, J. & Ruderman, M. A. in *I.A.U. Symp. No. 53: Physics of Dense Matter* (ed. Hansen, C. J.) 189-207 (Reidel, Dordrecht, 1974).
12. Ruderman, M. *Nature* **223**, 597-598 (1969).
13. Pines, D., Shaham, J. & Ruderman, M. A. *Nature phys. Sci.* **237**, 83, 96 (1972).
14. Alpar, M. A., Anderson, P. W., Pines, D. & Shaham, J. *Astrophys. J.* **276**, 325-334 (1984).
15. Pines, D. & Alpar, M. A. *Nature* **316**, 27-32 (1985).
16. Alpar, M. A., Cheng, K. S., Pines, D. & Shaham, J. *Mon. Not. R. astr. Soc.* **233**, 25-31 (1988).
17. Alpar, M. A., Cheng, K. S. & Pines, D. *Astrophys. J.* **346**, 823-832 (1989).
18. Manchester, R. N. & Peterson, B. A. *Astrophys. J.* **342**, L23-L25 (1989).
19. Manchester, R. N., Durdin, J. M. & Newton, L. M. *Nature* **313**, 374-376 (1985).

Probing Titan's atmosphere by stellar occultation

B. Sicardy*†, A. Brahic*†, C. Ferrari†, D. Gautier†, J. Lecacheux†, E. Lellouch†, F. Roques†, J. E. Arlot‡, F. Colas‡, W. Thuillot‡, F. Sèvre§, J. L. Vidal||, C. Blanco¶, S. Cristaldi¶, C. Buil#, A. Klotz# & E. Thouvenot#

* Université Paris 7, 2 place Jussieu, 75005 Paris, France

† Observatoire de Paris, 92195 Meudon Cédex Principal, France

‡ Bureau des Longitudes, 77 Avenue Denfert-Rochereau, 75014 Paris, France

§ Institut d'Astrophysique de Paris, 98 Avenue Arago, 75014 Paris, France

|| Observatoire du Pic du Midi et de Toulouse, 65200 Bagnères de Bigorre, France

¶ Istituto di Astronomia-Università di Catania, Viale Andrea Doria, 6, 95125 Catania, Italy

Observatoire Midi-Pyrénées, 14 Avenue Edouard Belin, 31400 Toulouse, France

WE report results from the first stellar occultation by Titan ever observed. As predicted by Wasserman¹, on 3 July 1989 the bright star 28 Sagittarii (visual magnitude, $V \approx 5.5$), passed behind Saturn's giant moon ($V \approx 8.3$), which is the only body in the Solar System that, like the Earth, has a dense, nitrogen-rich atmosphere². The event, visible from Europe, North Africa and the Middle East, allowed us to probe Titan's atmosphere in an altitude range of ~250-500 km (a pressure range of ~250-1 μ bar), where until now, there has been an 'information gap' between infrared and ultraviolet Voyager observations³⁻⁵. We also detected a central flash as the centre of Titan's shadow passed at a few tens of kilometres from Paris. This central flash allows us to estimate a finite oblateness of Titan's stratosphere, which could arise from a super-rotation of Titan's atmosphere.

We obtained the data presented here at the Paris Observatory at Meudon, the Pic du Midi Observatory in the French Pyrénées and the Catania Observatory on Mount Etna in Italy (see Table 1 for the instrument and geographic details). An accompanying paper by Hubbard *et al.*⁶ describes the results obtained in Israel

TABLE 1 Conditions of observations

Site	Telescope	Detector	Δt^*	Band	Longitude	Latitude
Meudon	1-m	SIT† vidicon	0.215	4,500–8,500 Å	02°13'53" E	48°48'18" N
	0.6-m	Fast photometer	0.0549	B		
Pic du Midi	2-m	Two-channel AsGa	0.05	8,900 ± 25 Å	00°08'32" E	42°56'12" N
		Fast photometer	—	7,500 ± 25 Å		
	1-m	SIT vidicon	0.561	B		
Catania	0.6-m	CCD‡	1.723	6,500 ± 200 Å		
	0.91-m	Fast photometer	1.0	B	14°58'42" E	37°41'30" N
	0.61-m	Three-channel	1.4	U, B, V		
		Simultaneous Photometer				

* Δt is the sample period (s).

† SIT, silicon intensified target.

‡ CCD, charge-coupled device.

and Italy. We determined the occultation chord at each station using the 'half-intensity' times t , that is, the times when the fraction of the unocculted stellar light hitting the Earth was 0.5. These times are found by fitting a synthetic light curve to the data, assuming that the atmosphere is isothermal, non-turbulent and transparent. These fits also provide the scale-height (the e-folding distance for the density as a function of altitude) H at the suboccultation points. The quantities t and H obtained at each site are given in Table 2, where H has been corrected for the limb curvature and gradient of the gravitational potential⁷.

Before connecting the suboccultation points by a circle, we have allowed small shifts in time to account for errors in absolute time, ~ 0.2 s from site to site. This procedure may cancel out small systematic effects due to Titan's atmospheric oblateness. A complete analysis of the satellite's shape will require all the available observations. The circle that best fits the positions of the suboccultation points has a radius of $2,977 \pm 5$ km, not including the data with saturation problems (see below). This radius must be corrected for the bending of rays due to refraction in Titan's atmosphere, which amounts to one scale-height H for an isothermal atmosphere. Relativistic bending due to Titan's mass (~ 0.2 km) is negligible here. Adopting $H = 54 \pm 6$ km (see below), we obtain a half-intensity radius of $R_{1/2} = 3,031 \pm 10$ km, corresponding to an altitude of 456 ± 10 km above Titan's surface ($R_s = 2,575$ km, ref. 3). The paths of 28 Sagittarii (28 Sgr), as observed from the different stations, are sketched in Fig. 3 of the accompanying paper⁶.

The average scale-height obtained at Pic du Midi is 51.9 ± 4 km at ingress and 59.1 ± 5 km at egress. At Catania, we obtain respectively 48.9 ± 3 km and 56.6 ± 3 km. We calculated the error bars from the variation in the scale-heights obtained at the same site; they give an idea of the internal consistency of the isothermal fit method. From an *a posteriori* analysis we found that the unocculted stellar flux recorded at the Meudon 1-m telescope and, to a lesser extent, in the 8,900-Å channel of the Pic du Midi 2-m telescope, were affected by saturation of the detector elements. Consequently, the parameters derived from those observations cannot be fully trusted, until we have done a complete calibration. Everything considered, we adopt a conservative average value of $H = 54.0 \pm 6$ km for the scale-height. We note that both our scale-height and half-light radius agree quite well with the results obtained by Hubbard *et al.*⁶— $H = 55.6 \pm 3.2$ km and $R_{1/2} = 3,031 \pm 6$ km.

The scale-height H yields the mean temperature $\bar{T}$ through $H = R\bar{T}/\bar{\mu}g$, where R is the gas constant, $\bar{\mu}$ is the mean molecular weight and g is the local gravity. From the mass of Titan (1.3457×10^{26} kg, ref. 8) we obtain $g = 97.7$ cm s⁻² at $R_{1/2} = 3,031$ km, and from $H = 54.0 \pm 6$ km we derive $\bar{T}/\bar{\mu} = 6.30 \pm 0.75$ K g⁻¹. From the recent re-analysis of the Voyager observations⁹ $\bar{\mu}$ is in the range 27.8–29.4 a.m.u. We note that only nitrogen and methane have been positively identified as

major components of Titan's atmosphere. A value of $\bar{\mu}$ greater than 27.95 a.m.u. would betray the presence of a heavier component, assumed to be argon on the basis of cosmological considerations¹⁰. Thus the possible range of $\bar{T}$ at the 456-km altitude level is $154 < \bar{T} < 207$ K. Assuming the absence of argon, we have $27.8 < \bar{\mu} < 27.95$ a.m.u., and thus $154 < \bar{T} < 197$ K, with a preferred value $T = 6.3 \times 27.8 = 175$ K.

The number density $n_{1/2}$ of the mid-occultation level is derived from the gas refractivity and the distance of the occulting body

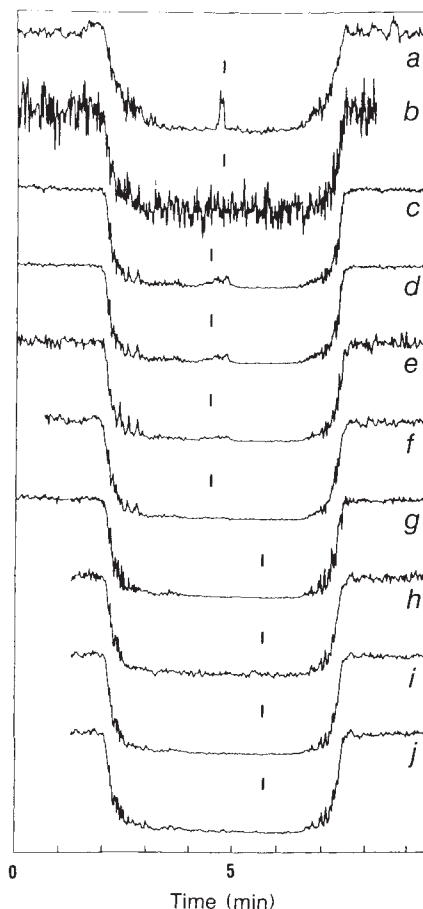


FIG. 1 The occultation light curves observed at Meudon (a, 1-m telescope; b, 60-cm telescope), Pic du Midi (c and d, 2-m telescope, 8,900 and 7,500 Å, respectively; e, 60-cm telescope; f, 1-m telescope) and Catania (g, 91-cm telescope; h, i, j, 61-cm telescope, U, B, V filters, respectively). See Table 1 for more details. The curves have been moved with respect to each other to fit in the same frame. The vertical tick marks on each curve indicate the time 22:42 (UT). The time scale has an arbitrary origin.

to the observer¹¹. The widest range of temperature derived above gives $n_{1/2}$ in the range $1.6\text{--}2.3 \times 10^{14} \text{ cm}^{-3}$, and a pressure range $3.4\text{--}5.1 \mu\text{bar}$ at the mid-occultation level. These values may be compared to the model of Lellouch *et al.*¹², which predicts a temperature of 174 K and a pressure of $4.5 \mu\text{bar}$ at an altitude of 450 km. Considering the uncertainties of the model, the agreement with the present observation is quite satisfactory.

We use the flash profiles observed at Meudon and Pic du Midi (Figs 1 and 2) to estimate the oblateness of Titan's stratosphere and the absorption at visible wavelengths in this region of the atmosphere. We used, in a preliminary approach, a model on an isothermal and non-turbulent atmosphere, previously applied to Mars' and Neptune's atmospheres^{13–15}. Thus, if we assume that the central flash at Meudon is not distorted by turbulence, then the double-peaked structure implies that the path of this observatory in Titan's shadow crossed Titan's evolute, and is therefore inconsistent with a perfectly spherical atmosphere. Our model best matches the observed structure for an apparent oblateness of 0.014 at the density level $n = 10^{16} \text{ cm}^{-3}$ (that is, at an altitude of 250 km, at $\sim 0.25 \text{ mbar}$) and an atmospheric extinction of $\sim 90\%$ for an atmospheric path of 0.35 km atm. Assuming a titanocentric elevation of the Earth $B = 25.4^\circ$, the real oblateness is 0.017. The fit also indicates that Titan's shadow centre passed 20 km south of Meudon, in the plane of the sky. This corresponds to a distance of $\sim 60 \text{ km}$ south of Paris at the surface of the Earth, once the elevation of Titan above the horizon is taken into account.

Because the Pic du Midi trajectory passed at $\sim 190 \text{ km}$ from Titan's shadow centre (in the plane of the sky), the flash observed there is to a large extent insensitive to Titan's atmospheric oblateness. The complex structure is conceivably due to atmospheric turbulence effects¹⁶. Finally, a preliminary analysis of the central flash observed at Manley Observatory (UK) (R. Miles, personal communication) shows that the flash profile is consistent with the oblateness derived at Meudon and, in

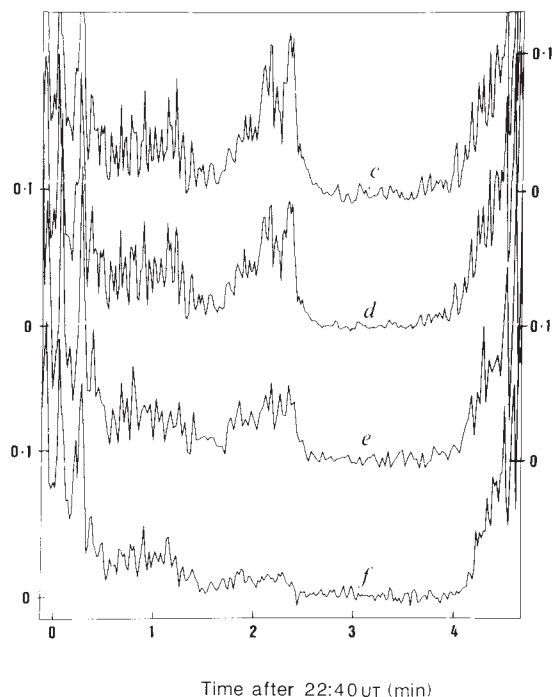


FIG. 2 Vertically expanded view of the Pic du Midi central flash. The vertical scale is relative to the unocculted stellar flux. *c*, 8,900-Å channel of the 2-m telescope; *d*, 7,500-Å channel of the 2-m telescope; *e*, 60-cm telescope (6,500 Å); *f*, 1-m telescope (4,500 Å). Note the strong wavelength dependence of the flash absorption, which is not observed for the spikes. Note also that the stellar flux is more absorbed during the second half of the occultation.

TABLE 2 Results of the isothermal fits

Site	Telescope	t_i (UT)	H_i (km)	t_e (UT)	H_e (km)
Meudon	1-m	22:39:26.2	(87.4)	22:44:28.7	(100.1)
	0.6-m	22:39:20.1	(78.9)	22:44:36.5	(54.9)
Pic du Midi	2-m	22:39:36.8	(47.8)	22:44:55.3	(46.7)
		22:39:36.7	51.2	22:44:54.8	52.9
	1-m	22:39:38.2	57.8	22:44:53.0	60.8
	0.6-m	22:39:37.4	46.9	22:44:55.9	63.5
Catania	0.91-m	22:38:42.2	51.7	22:43:52.5	60.8
	0.61-m U	22:38:44.0	52.4	22:43:52.6	58.3
	0.61-m B	22:38:42.9	44.6	22:43:53.6	52.1
	0.61-m V	22:38:43.0	46.7	22:43:53.1	55.0

The quantities t and H are the half-light time and the scale-height, respectively, and the subscripts *i* and *e* refer to ingress and egress of the star, respectively. Numbers in parentheses are uncertain because of calibration problems (see text).

particular, that the central flash near the evolute is little affected by turbulence in Titan's atmosphere. Also, the timing of the Manley central flash constrains Titan's pole orientation to be within $\sim 2.5^\circ$ of Saturn's pole.

If real, the atmospheric oblateness could be due to a super-rotation of the atmosphere or a differential thermal effect. If entirely because of rotation, a 1.7% oblateness would require a minimum 26-h rotation period (assuming a gravitational quadrupole moment $J_2 = 0$). If $J_2 > 0$, the period may be larger than 26 h (ref. 15). In fact, both the evaluation of the thermal winds from the Voyager infrared observations and the processing of a large number of images from the spacecraft, indicate a super-rotation of the stratosphere with respect to the solid body of Titan (assumed to rotate with the synchronous period of $\sim 16 \text{ d}$) by a factor of 5–10 (refs 17, 18). If this interpretation is correct, then Titan's large-scale stratospheric motions may be compared to those of Venus. If purely thermal, the apparent oblateness would correspond to an integrated difference of 50 km, that is, one scale-height, between the altitudes of the equatorial and the polar 0.25-mbar isobars. This is unlikely if the equator-to-pole contrast is as low as 20 K at $\sim 200 \text{ km}$, as observed at the time of the Voyager encounter^{19,20}.

Figure 2 shows that the atmospheric extinction at the 250-km level is highly wavelength dependent, with stronger absorption towards shorter wavelengths. The central flash at 8,900 Å may be affected by the methane absorption band, whereas the 4,500 Å flash is barely above the noise. From the ratio of the flashes at 7,500 Å and 6,500 Å the extinction seems to vary as $\sim \lambda^4$, indicating that Rayleigh scattering is responsible, at least in part, for the extinction of the central flash. We also note (Fig. 2) that the spikes observed during the first half of the occultation (that is, higher in the stratosphere) are much less affected by this wavelength-dependent extinction.

The overall occultation profiles at Pic du Midi and Catania have an asymmetric shape (Figs 1 and 2). Whereas 28 Sgr was clearly detected during the first half of the occultation, only faint spikes are visible during the second half in the 8,900-Å and 7,500-Å channels of Pic du Midi. This extinction may be due to an absorbing aerosol layer and/or a colder layer in the altitude range 250–450 km. Note that the star first shone through the evening limb in the southern, winter, hemisphere of Titan, and is then absorbed in the morning equatorial limb. A careful analysis of all the light curves should be undertaken to define better the absorbing properties and the location of that layer, which in turn may give some constraints on the seasonal and diurnal variations in the satellite's atmosphere.

A comparison of all the data obtained by the different groups is needed to improve the values given here. In particular, a comparative analysis of the central flashes reported in Northern

Europe could be fruitful. Also, a more complete analysis requires the inversion of the light curves, rather than isothermal fits to the data. The derived temperature at a pressure of $\sim 3\text{--}5\ \mu\text{bar}$ is consistent with present aeronomy models based on Voyager observations. Diurnal and/or seasonal variations, however, could explain some of the observed differences. Finally, predictions for future occultations by Titan followed by their observation would place further constraints on models of Titan's stratosphere before the Cassini mission starts (in 2003) its exploration of this unique satellite. $\square$

Received 14 September; accepted 27 November 1989.

1. Wassermann, L. H. *IAU Circ.* No. 4761 (1989).
2. Hunten, D. *et al.* in *Saturn* (eds Gehrels, T. & Matthews, M. S.) 671–759 (University of Arizona Press, 1984).
3. Lindal, G. F. *et al.* *Icarus* **53**, 348–363 (1983).
4. Hanel, R. A. *et al.* *Science* **212**, 192–200 (1981).
5. Broadfoot, A. L. *et al.* *Science* **212**, 206–211 (1981).
6. Hubbard, W. B. *et al.* *Nature* **343**, 353–355 (1990).
7. Elliot, J. L. *et al.* *Icarus* **77**, 148–170 (1989).
8. Burns, J. in *Satellites* (eds Burns, J. & Matthews, M. S.) Ch. 1 (University of Arizona Press, 1986).
9. Lellouch, E. *et al.* *Icarus* **79**, 328–349 (1989).
10. Owen, T. C. *Planet. Space Sci.* **30**, 833–838 (1982).
11. Elliot, J. L. *A. Rev. Astr. Astrophys.* **17**, 445–475 (1979).
12. Lellouch, E., Hunten, D. M., Kockarts, G. & Coustenis, A. *Icarus* (in the press).
13. Elliot, J. E. *et al.* *Astrophys. J.* **217**, 661–679 (1977).
14. Lellouch, E., Hubbard, W. B., Sicardy, B., Vilas, F. & Bouchet, P. *Nature* **324**, 227–231 (1986).
15. Hubbard, W. B. *et al.* *Icarus* **72**, 635–646 (1987).
16. Hubbard, W. B. *et al.* *Astrophys. J.* **325**, 490–502 (1988).
17. Conrath, B. J. in *The Atmospheres of Saturn and Titan*, *ESA Sci. Publ.* 241 (eds Roife, E. & Battrick, B.) 83–94 (European Space Agency, 1985).
18. Wenkert, D. D. & Garneau, G. V. *Bull. Am. astr. Soc.* **19**, 875 (1987).
19. Flasar, F. M., Samuelson, R. E. & Conrath, B. J. *Nature* **292**, 693–698 (1981).
20. Coustenis, A., Bézard, B. & Gautier, D. *Bull. Am. astr. Soc.* **21**, 959 (1989).

ACKNOWLEDGEMENTS. We thank J. L. Leroy, who generously gave two nights of his own observing time for this experiment, C. Hubert and S. Pau for their efforts to ensure the preparation of one of the telescopes, R. Miles for access to his data before publication, and R. G. French, D. M. Hunten and T. Owen for discussions and comments.

Results for Titan's atmosphere from its occultation of 28 Sagittarii

W. B. Hubbard*, **D. M. Hunten***, **H. J. Reitsema†**,
N. Brosch‡, **Y. Nevo§**, **E. Carreira||**, **F. Rossi||**
& **L. H. Wasserman¶**

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Ball Aerospace Systems Division, Boulder, Colorado 80306, USA

‡ Wise Observatory, Tel Aviv University, Ramat Aviv, Tel Aviv 69978, Israel

§ Kibbutz Ein Harod, Meuchad 18965, Israel

|| Specola Vaticana, Vatican Observatory Research Group, I-00120 Città del Vaticano

¶ Lowell Observatory, Flagstaff, Arizona 86001, USA

ON 3 July 1989 the bright K giant star 28 Sgr was occulted by Saturn's largest moon, Titan. This event, which was predicted by Wasserman¹, offered a unique opportunity to probe Titan's extensive nitrogen-rich atmosphere in an altitude range not investigated by the Voyager 1 spacecraft^{2,3}. Our group observed the occultation from three stations in the Mediterranean area, and here we examine the data set. We derive average mesospheric temperatures of $\sim 180\text{ K}$, with evidence for lateral and vertical atmospheric inhomogeneities on scales ranging from $\sim 10\text{--}1,000\text{ km}$. Our results are consistent with published models⁴ of Titan's mesosphere.

Table 1a gives the locations and wavelengths of observation for our three stations. Two of the stations (the Vatican Observatory near Rome and the Wise Observatory in southern Israel)



FIG. 1 Ground track for the occultation of 28 Sgr by Titan (top to bottom: northern limit, centre line, southern limit).

were at established telescope sites. The third station, at Kibbutz Ein Harod in northern Israel, was a portable telescope situated to investigate the homogeneity of Titan's atmosphere on a scale of $\sim 130\text{ km}$ by comparison with data from the Wise Observatory. The deployment of our experiments generally in the southern part of Titan's shadow was coordinated with Sicardy *et al.*⁵, who observed generally to the north. All investigators tried to secure as wide coverage as possible consistent with geographical and logistic limitations. Figure 1 shows the ground track for the occultation, which was deduced from the timings reported below.

The instruments at the Vatican Observatory and Ein Harod were identical 100-Hz two-channel occultation photometers⁶. The photometer⁷ at the Wise Observatory was operated at 20 Hz and was also a two-channel experiment but the second channel monitored the (constant) signal from a nearby comparison star. The Titan channel was equipped with a filter of full width at half maximum $0.050\ \mu\text{m}$ centred at $0.730\ \mu\text{m}$. In the blue channels ($0.45\ \mu\text{m}$) at the Vatican Observatory and Ein Harod, the

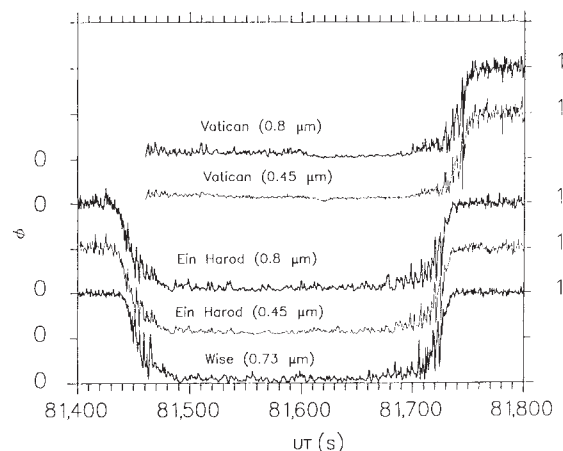


FIG. 2 Occultation light curves, smoothed to a time resolution of 0.5 s. The Vatican curves have been shifted by 100 s.

TABLE 1 Parameters

(a) Station					
Location	Telescope	λ (μm)	Latitude	Longitude	Elevation (m)
Vatican Observatory	61-cm	0.45, 0.8	41°44'48" N	12°39'6" E	450
Ein Harod	36-cm	0.45, 0.8	32°33'35" N	35°23'42" E	20
Wise Observatory	1-m	0.73	30°35'48" N	34°45'48" E	900

(b) Light curve				
Location	$t_{1/2}(\text{red})$ (UT)	$t_{1/2}(\text{blue})$ (UT)	$H(\text{red})$ (km)	$H(\text{blue})$ (km)
Vatican Observatory (egress)	81,842.1 $\pm$ 0.5	81,842.7 $\pm$ 0.5	57 $\pm$ 2	57 $\pm$ 2
Ein Harod (ingress)	81,447.3 $\pm$ 0.5	81,448.2 $\pm$ 0.5	59 $\pm$ 2	60 $\pm$ 2
Ein Harod (egress)	81,724.4 $\pm$ 0.5	81,742.2 $\pm$ 0.5	52 $\pm$ 2	54 $\pm$ 2
Wise Observatory (ingress)	81,453.4 $\pm$ 0.5		56 $\pm$ 2	
Wise Observatory (egress)	81,723.6 $\pm$ 0.5		50 $\pm$ 2	
			mean $H = 55.6 \pm 3.2$ km	

unocculted stellar signal was ~15 times the background signal of Titan plus sky, whereas in the red channels (0.8 μm ; 0.73 μm) at all three stations the signal/background ratio was ~30.

The data acquisition systems at the Vatican Observatory and Ein Harod were synchronized to coordinated Universal Time (UT) before their departure from Arizona. We synchronized the timing at the Wise Observatory to a Soviet radio time service, and we compared the timing of the Ein Harod and Wise instruments on the day after the event, which facilitated the detection and correction of an unexplained slippage by an integral number of seconds in the Ein Harod system. The precision of the system times is at least as good as ± 0.1 s.

The conditions for both of the stations in Israel were photometric. Heavy clouds were initially present during the ingress at the Vatican Observatory, but cleared shortly thereafter, and we obtained good data for the rest of the event. Figure 2 shows the data sets obtained at the three stations, where ϕ is the stellar flux in units of the unocculted flux. For convenience, the data from the Vatican Observatory have been shifted to earlier times by exactly 100 s.

From calibration data obtained for the star and Titan separately, we find that the minimum normalized stellar flux ϕ in the middle of the occultation for the two stations in Israel was ~0.05. Thus, the fluctuations in the light curves during the occultation are significant—the star remained visible throughout the occultation. Apparently the brightest refracted image moved all the way around the southern limb of Titan without suffering absorption, as there is no clear evidence from preliminary reduction of the data that the minimum value of ϕ in the data from observations in Israel depends on wavelength.

The light curve observed from the Vatican is different, with a significant reduction in the signal and smoothing of the fluctuations during the second half of the occultation. The quenching

of fluctuations in the stellar signal is more apparent in the blue channel. Although we were initially skeptical of this behaviour because of the presence of (terrestrial) clouds during ingress, it correlates well with data reported by Sicardy *et al.*⁵ from nearby chords, and therefore seems to be real.

We deduced the geometry of the occultation from the times of half-flux ($t_{1/2}$) obtained by fitting Baum-Code light curves⁸ to 40 s of data centred on ingress or egress. This procedure also yielded the best-fit value of the scale-height H at this level in Titan's atmosphere. Results from the fits, along with estimated errors, are given in Table 1b. Our mean scale-height, 56 km, can be compared with the 54 km reported in ref. 5. There is no significant difference at the present state of the analyses.

Figure 3 shows the geometry deduced from our half-flux times and those reported by Sicardy *et al.* In Fig. 3, celestial north is up and celestial east is to the left. Ingress points are on the right and egress points are on the left. Dots show our half-flux timings and crosses show the timings of Sicardy *et al.* The heavy solid circle shows the outline of Titan's surface, at a radius³ of 2,575 km. The lighter solid circle shows our best-fit half-flux shadow locus, with a radius of $2,975 \pm 5$ km. We obtain the corresponding radius in Titan's atmosphere by adding a refraction correction, which equals one scale-height; thus the true radius is $3,031 \pm 6$ km, and the altitude above Titan's surface $z = 456 \pm 6$ km. Our data fill a gap in the information obtained^{2,3} by Voyager 1. The Voyager radio-occultation experiment³ obtained data on Titan's neutral atmosphere from $z = 0$ to $z = 200$ km (circle with short dashes in Fig. 3), whereas the Voyager ultraviolet-spectrometer experiment² measured the atmosphere in the vicinity of $z = 1,270$ km (circle with long dashes).

To obtain a temperature from the observed scale-height, we must assume a mean molecular mass, which was taken to be exactly 28 a.m.u., following the analysis of Voyager data³. The

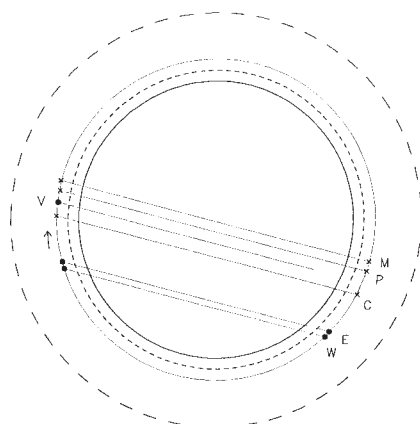


FIG. 3 Location of occultation points on Titan's limb (M=Meudon, P=Pic du Midi, C=Catania, E=Ein Harod, W=Wise, V=Vatican). Time increases right to left (opposite to Figs 2 and 4). Reading outward, the circles represent the solid surface, the limit of Voyager infrared and radio data, our half-flux level, and the region measured in Voyager ultraviolet occultation.

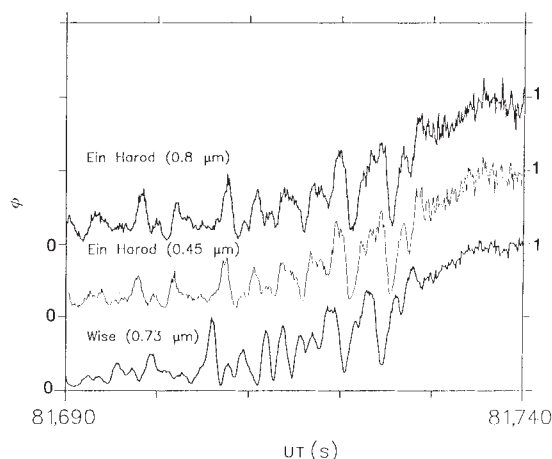


FIG. 4 High-resolution view of emersion at the stations in Israel.

composition could be pure nitrogen or a mixture including small amounts of methane and argon³. With gravity $g = 97.7 \text{ cm s}^{-2}$, our mean scale-height corresponds to a temperature T of $183 \pm 11 \text{ K}$ and a pressure P of $5.1 \pm 0.3 \mu\text{bar}$. These values agree reasonably well with the model of Lellouch *et al.*⁴, which gives $T = 174 \text{ K}$, $P = 4.5 \mu\text{bar}$ at $z = 450 \text{ km}$.

There is evidence that Titan's atmosphere is not spherically symmetric. First, there is a weak indication (Table 1b) that the scale-heights are systematically slightly larger on Titan's ingress (sunset) limb compared with those on the egress (sunrise) limb. Second, the data taken close to a central chord (Vatican) show evidence of suppression of the stellar flux, but only when the stellar image moves toward the eastern (sunrise) limb (arrow in Fig. 3). There is no clear sign of such asymmetric suppression in the data from Israel, where the stellar image would have stayed slightly higher in Titan's atmosphere and not moved as far north. A detailed combined analysis of all the occultation data should yield a map of absorption and/or temperature gradients in Titan's mesosphere.

On smaller scales, we see evidence of scintillations caused by highly anisotropic density fluctuations in Titan's upper atmosphere. Figure 4 shows an expanded view of the egress observed from the two stations in Israel. These stations were separated by 132 km (more than two scale-heights) along Titan's limb at egress, yet they show detailed correlation of fluctuations once the stellar image has emerged beyond the mean half-flux point. There is even an indication (Fig. 2) of some correlation between the red channels throughout the event. The two-channel data from Ein Harod show effects of dispersion in Titan's atmosphere (the blue scintillations tend to occur earlier than the red scintillations during egress, and later during ingress). This effect, which is also visible in the data from the Vatican, can be used to establish an independent density scale for detailed studies of Titan's atmospheric structure. □

Received 21 September; accepted 28 November 1989.

1. Wasserman, L. H. *IAU Circ. No. 4761* (1989).
2. Smith, G. R. *et al. J. geophys. Res.* **87**, 1351–1359 (1982).
3. Lindal, G. F. *et al. Icarus* **53**, 348–363.
4. Lellouch, E., Hunten, D. M., Kockarts, G. & Coustenis, A. *Icarus* (in the press).
5. Sicardy, B. *et al. Nature* **343**, 350–353 (1990).
6. Hubbard, W. B. *et al. Astr. J.* **90**, 655–667 (1985).
7. Vidal, N. V., Brosch, N. & Livio, M. *Observatory* **98**, 60–62 (1978).
8. Baum, W. A. & Code, A. D. *Astr. J.* **58**, 108–112 (1953).

ACKNOWLEDGEMENTS. We thank G. Coyne and S. Navon for help in arranging the Vatican and Ein Harod experiments. This work has been supported by NASA and, at the Wise Observatory, by grants from the Israeli Academy of Sciences and the Israeli Ministry of Science and Development.

Imaging of spatio-temporal pattern evolution during carbon monoxide oxidation on platinum

H. H. Rotermund, W. Engel, M. Kordesch* & G. Ertl

Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4–6, D-1000 Berlin 33, FRG

* Present address: Department of Physics, University of Ohio, Athens, Ohio 45701–2979, USA

THE rates of chemical reactions may exhibit temporal oscillations, even when the reaction parameters are kept constant. These oscillations might be associated with spatial pattern formation owing to local variations of concentrations or temperature, or both, which propagate as chemical waves^{1,2}. An example is the catalytic oxidation of carbon monoxide on the surfaces of platinum single crystals. The mechanism of this reaction is well known, and under isothermal, low-pressure conditions, many types of temporal self-organization, ranging from harmonic oscillations to chaotic behaviour, have been observed^{3–6}. Spatial patterns have also been demonstrated by scanning beams of either electrons³ or ultraviolet photons⁶ across the surface, and by monitoring the electrons reflected or photoemitted, respectively. However, both scanning techniques are hampered by limited temporal as well as lateral resolution. Here we present observations of an unexpectedly rich variety of self-sustained spatio-temporal patterns by continuous imaging (on a timescale of $\geq 1 \text{ ms}$) of the surfaces using a newly developed, photoemission electron microscope⁷ with a spatial resolution of a few micrometres. Pattern evolution can be followed on a much finer timescale than is available from other imaging techniques. We observe a new type of structure with strong similarities to those found in the Belousov–Zhabotinsky oscillatory reaction.

The catalytic formation of CO_2 from CO and O_2 on platinum proceeds by recombination of chemisorbed CO and O species, the latter originating from dissociative adsorption of O_2 . Under

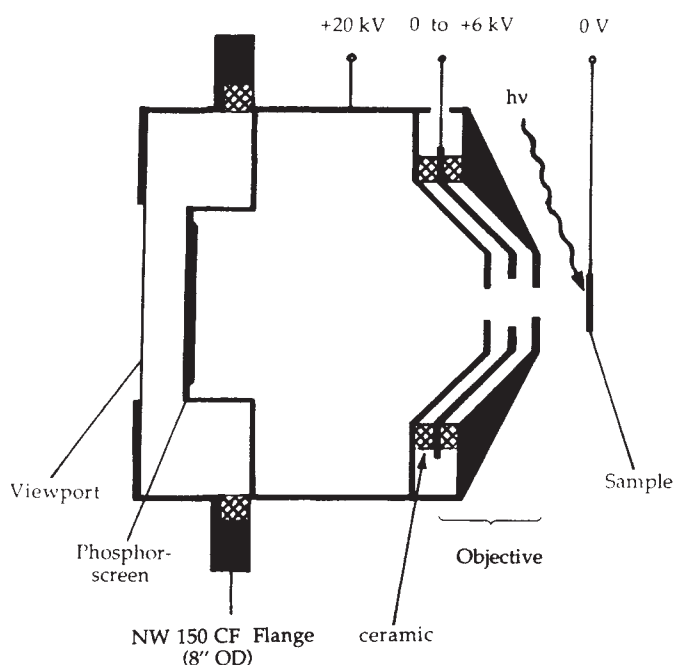


FIG. 1 Schematic cross-section of the photoemission electron microscope used for imaging of spatial pattern formation on surfaces.

low-pressure ($<10^{-3}$ mbar) conditions and for certain sets of control parameters (CO and O₂ partial pressures, temperature), kinetic oscillations are observed on Pt(100) and Pt(110) surfaces which are linked with periodic, adsorbate-induced transformations of the atomic surface structure³⁻⁵. Essentially, the surface switches locally between states of high CO and O coverage, and spatial coupling occurs either through wave propagation initiated by surface diffusion/reaction or (practically instantaneously) through the gas phase (that is, by small variations in partial pressure of the reactant gases). Temporal oscillations on Pt(100) are usually rather irregular and often exhibit fairly long periods with durations of the order of minutes. It was for these conditions that scanning techniques^{3,6} (with image-recording times >10 s) have provided information about spontaneous pattern formation and propagation of concentration waves across the surface. Whether similar features are associated with the more rapid oscillations that also occur with this system remained unclear, however, because of the limited time resolution. With the Pt(110) surface, on the other hand, the oscillations are generally regular and occur with periods of the order of seconds. On the basis of indirect evidence, it was concluded⁵ that different regions on the surface couple to each other through the gas phase, but no direct information about the formation of spatial patterns could be obtained. Such information is, however, of particular interest, because the system may exhibit a well defined transition from regular oscillations to chaos through a sequence of period-doublings⁴.

The state of the surface can be sensitively monitored through the work function ϕ which, for example, increases on a Pt(100) surface by ~ 0.3 eV in changing from coverage with adsorbed CO to adsorbed O. The yield of electrons emitted from a surface irradiated by ultraviolet photons (with energy $h\nu$) close to the threshold frequency is approximately proportional to $(e\phi - h\nu)^2$ (Fowler equation) and can therefore conveniently be used for imaging local variations of the surface composition. This is the

basis of the scanning photoemission microscope (SPM)⁶ as well as the photoemission electron microscope (PEEM). The latter has been used previously for investigation of the Pt(100)/CO + O₂ system⁷, but in that study the conditions were outside the region of sustained oscillatory behaviour.

The PEEM as used in this study is illustrated in Fig. 1. The sample surface is irradiated with ultraviolet light from a deuterium discharge lamp with a broad spectral distribution up to $h\nu = 6.5$ eV, centred around 6.0 eV. The photoemitted electrons are accelerated to 20 keV over a distance of 4 mm to the front electrode of an electrostatic three-electrode lens⁸ which creates an image with about tenfold magnification on a 'backview' fluorescence screen. The sample is at ground potential, and the two outer electrodes of the lens, the microscope tube and the fluorescent screen are at high voltage. The potential of the centre electrode is close to ground and may be varied for focusing the image. With the present instrument, only the central region is imaged distortion-free. A more elaborate three-lens system now under construction will partly overcome this disadvantage and also provide far higher magnifications. The device is mounted on an 8 inch stainless steel flange with a viewport and is attached to a standard ultra-high-vacuum system⁶. The image on the fluorescent screen can be photographed directly with a time constant down to 1 ms, or recorded on video tape by use of a CCD camera. In the latter case images are collected with a repetition frequency of 50 Hz, corresponding to a time resolution of 25 ms, which is about three orders of magnitude faster than with the SPM⁶ or scanning LEED³ techniques.

Kinetic oscillations were initiated with sample temperatures typically between 380 and 530 K and O₂ and CO partial pressures of up to several times 10^{-4} mbar; these conditions create no distortion of the electron optical imaging process, and the high electric field in the presence of the gas atmosphere did not affect the surface properties. The laterally integrated temporal behaviour of the surface reaction can be monitored most

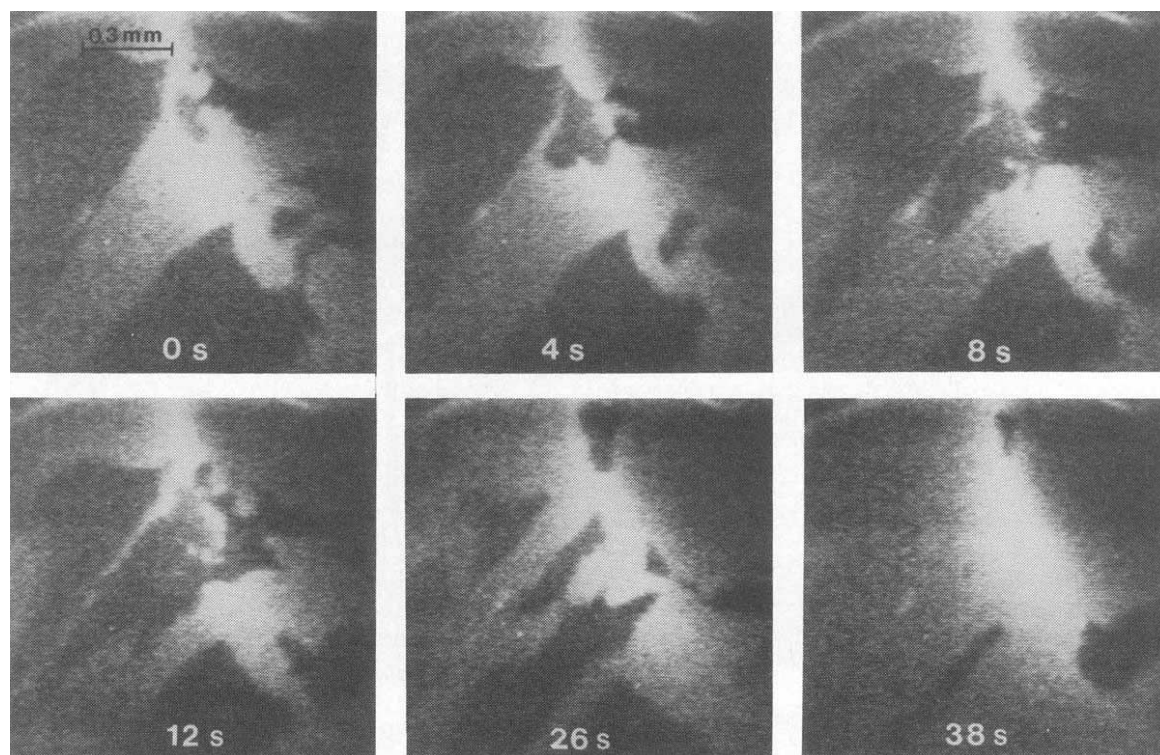


FIG. 2 A series of spatial patterns on the Pt(100) surface accompanying sustained kinetic oscillations during catalytic oxidation of carbon monoxide. Experimental conditions: temperature = 480 K, partial pressures $p_{\text{O}_2} =$

2.85×10^{-4} mbar, $p_{\text{CO}} = 2.0 \times 10^{-5}$ mbar. The images were photographed from the video screen with illumination times of 0.2 s.

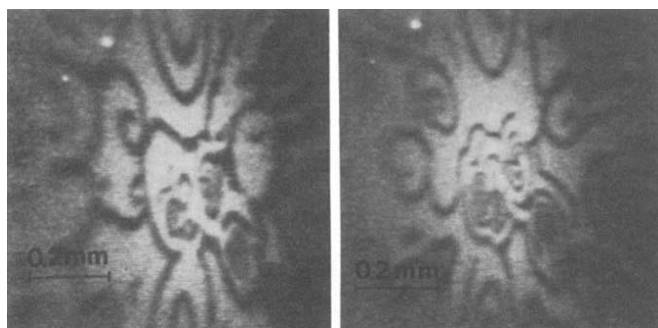


FIG. 3 Spatial pattern evolution during oscillatory CO oxidation on a Pt(110) surface. The two images were recorded 1 min apart. Temperature = 433 K, $p_{O_2} = 3.1 \times 10^{-4}$ mbar, $p_{CO} = 3.2 \times 10^{-5}$ mbar.

conveniently by recording the total current yield from photo-emitted electrons (typically of order 10^{-9} A) through the sample during ultraviolet irradiation.

A typical set of rapidly changing patterns associated with oscillatory CO oxidation on the Pt(100) surface is shown in Fig. 2. Dark areas, reflecting a high concentration of adsorbed O atoms and therefore a high work function, form spontaneously and propagate in an irregular manner across the surface. The velocity of propagation of these 'oxygen waves' $\sim 0.1 \text{ mm s}^{-1}$. A high O concentration leads to a high reaction rate. The fraction of this 'dark' surface reaches a maximum after about 15 s; the domains of adsorbed O then diminish to a minimum after 40 s. The general features are qualitatively similar to those observed previously with the same system under conditions of slower wave propagation and of longer timescales of the (integral) rate oscillations⁶; that is, spontaneous symmetry breaking of the surface composition at random locations followed by the propagation of chemical waves with irregular shapes. The highly improved resolution makes it possible, however, to follow much faster processes and to identify more subtle details in these patterns.

Sometimes fairly isotropic trigger waves of O islands were observed, starting from only a few nucleation centres and growing in roughly circular patterns on a timescale of several minutes. In these cases the oscillations have large amplitudes and fairly regular periodicity.

For the Pt(110) surface, regular high-frequency oscillations are often observed which have been tentatively attributed to the dominant effect of spatial coupling through the gas phase⁵. Indeed, on this surface the simultaneous appearance of a high density of domains of a new phase was frequently observed whose growth was anisotropic, being faster along the crystallographic [110] direction of the surface. Under different conditions, extended wave fronts propagated very rapidly ($\geq 1 \text{ mm s}^{-1}$) and periodically across the whole surface, corresponding to regular temporal oscillations with large amplitudes.

For particular sets of external parameters it was found that the apparently simultaneous formation (within $\ll 1$ s) of numerous nucleation centres was followed by their fairly slow evolution by wave propagation into patterns of the type shown in Fig. 3. In spite of the distortion of the outer parts of the images by the imperfections of the electron optics, the concentric character of these patterns is obvious and is reminiscent of features observed in three-dimensions in oscillatory chemical reactions such as the Belousov-Zhabotinsky reaction^{9,10}. To our knowledge, this is the first experimental observation of spiral waves in an initially uniform, truly two-dimensional system. closer inspection suggests that these patterns are self-repeating on varying length scales. Detailed analysis of such spatio-temporal patterns will be the aim of future systematic investigations. □

Received 6 November; accepted 12 December 1989.

1. Nocolis, G. & Baras, F. (eds) *Chemical Instabilities* (Reidel, Dordrecht, 1984).
2. Field, R. J. & Burger, M. (eds) *Oscillations and Travelling Waves in Chemical Systems* (Wiley, New York, 1985).
3. Imbihl, R., Cox, M. P. & Ertl, G. *J. chem. Phys.* **84**, 3519–3534 (1986).
4. Eiswirth, M., Krischer, K. & Ertl, G. *Surf. Sci.* **202**, 565–591 (1988).
5. Eiswirth, M., Möller, P., Wetzl, K., Imbihl, R. & Ertl, G. *J. chem. Phys.* **90**, 510–521 (1989).
6. Rotermund, H. H., Jakubith, S., Von Oertzen, A. & Ertl, G. *J. chem. Phys.* **91**, 4942–4948 (1989).
7. Kordesch, M. E., Engel, W., Lapeyre, G. J., Zeitler, E. & Bradshaw, A. M. *Appl. Phys.* **A49**, 399–402 (1989).
8. Griffith, O. H. & Rempfer, F. F. *Adv. opt. Electron Microsc.* **10**, 269–337 (1987).
9. Winfree, A. T. in *Oscillations and Travelling Waves in Chemical Systems* (eds Field, R. J. & Burger, M.) 441–472 (Wiley, New York, 1985).
10. Noszticzius, Z., Horsthemke, W., McCormick, W. D., Swinney, H. L. & Tam, W. Y. *Nature* **329**, 619–620 (1987).

ACKNOWLEDGEMENTS. We thank S. Jakubith, S. Kubala and A. von Oertzen for technical assistance.

Recent dyke-induced large-scale block movement at Mount Etna and potential slope failure

W. J. McGuire*, A. D. Pullen† & S. J. Saunders*

* Volcanic Studies Unit, Department of Human and Environmental Sciences, West London Institute, Isleworth, Middlesex TW7 5DU, UK

† Department of Civil Engineering, Imperial College, London SW7 2BP, UK

HERE we report the results of a geodetic survey undertaken on Mount Etna between 1981 and 1988. Observed horizontal ground movements in the vicinity of the Valle del Bove caldera, demonstrate that dyke emplacement during 1983 and 1985 resulted in the 2.8-m eastward displacement of a block of western caldera rim. A history of past instability and collapse along the rim is indicated by the presence of large back-rotated blocks of prehistoric volcanics¹. Furthermore, the unusually linear form of the present western rim indicates past slope failure along north-south aligned surfaces, and therefore that dyke-induced collapse may have been an important agent in caldera enlargement. We suggest a self-sustaining mechanism, whereby the stress regime generated as a consequence of the caldera's existence, results in the repeated injection of north-south orientated dykes adjacent to the western rim^{2,3}. Dilatation associated with magma intrusion leads in turn to periodic phases of slope instability and failure of the cliff wall, causing a progressive westward migration of the western rim.

A network of geodetic stations, established on Etna in 1971 (ref. 4), was extended during 1981 (Fig. 1) to record ground deformation changes in the upper part of the southern 'rift zone'.

TABLE 1 Cumulative group displacement behaviour, 1981–1987

Station group	Location	Direction	Displacement* Degree
A	West of 1983/85 fissures	West	Moderate Max. displacement 90 cm. Displacement decreasing away from the fissures
B	South and south-east of 1983/85 fissures	East and east-south-east	Small Max. displacement 50 cm
C	East of 1983/85 fissures	East and east-south-east	Considerable Max. displacement > 280 cm
D	Immediately south of south-east crater	South-south-west	Moderate Max. displacement 70 cm

* Relative to station 14 (see Fig. 3).

During 1981, we made 89 distance measurements using a Hewlett Packard HP3800B EDM (electronic distance meter), and a Kern DKM2-A 1-s theodolite. As well as establishing distances between stations, relative heights were measured to ± 5 cm. Redundancy within the network provided an estimate of the

accuracy of the distance measurements which, in the case of slope distances, was generally better than 5 mm (± 5 p.p.m.). The coordinates for each station were projected onto a sphere at sea level, giving an estimated maximum standard error of ~ 2.5 cm.

We partially resurveyed the geodetic network (17 lines) in April 1983 during the early stages of the 1983 eruption, and again (27 lines) in September of that year, following the end of the eruption. We selected lines that would provide the maximum information about movements related to the emplacement of the 1983 feeder dyke and the development of the associated fissure system. More complete reoccupation of the network has since taken place—in July 1985, towards the end of the 1985 eruption (21 lines), in July 1987 (68 lines) and in July 1988 (73 lines). Since the establishment of the network, we have used a range of equipment for re-measurement including EDM models AGA-12, AGA-14, Sokkisha RED 2L and Sokkisha RED 2A, and Kern DKM2 and DKM3 theodolites. The range of equipment used, together with the lack of redundancy resulting from only partial re-measurement of the network on some occasions, is far from ideal and may have resulted in errors as great as 5 cm in some horizontal distances. Given the scale of the movements recorded over this period (up to 2.8 m along some lines), however, we do not consider these errors to be significant.

The consecutive data sets show that significant horizontal movements are related to the emplacement of shallow feeder dykes associated with the March 1983 (ref. 5) and March 1985 eruptions. East-west lines crossing the 1983 and 1985 fissures display length increases of 1–2 m between 1981 and April 1983, and again between September 1983 and July 1985 (Fig. 2). By contrast, only relatively small length changes, ranging from +30 cm to –10 cm, are recorded along the same lines between April and September 1983 and between July 1985 and July 1987. Over the six-year period 1981–1987 a generally consistent framework of movement is apparent in which extension along an east-west axis is accompanied by a corresponding contraction in lines orientated north-south and parallel to the fissures. Length changes between 1981 and 1988 along three such lines (Fig. 2) show that shortening is progressive, has continued at roughly the same rate between 1981 and 1987, and amounts to a cumulative contraction of under 70 cm. Exceptional behaviour is demonstrated over the period 1985 to 1987 by lines (8–6, 8–5 and 8–7) that cross the fissures and extend to the edge of the Valle del Bove (Fig. 2). These lines continued to show increases

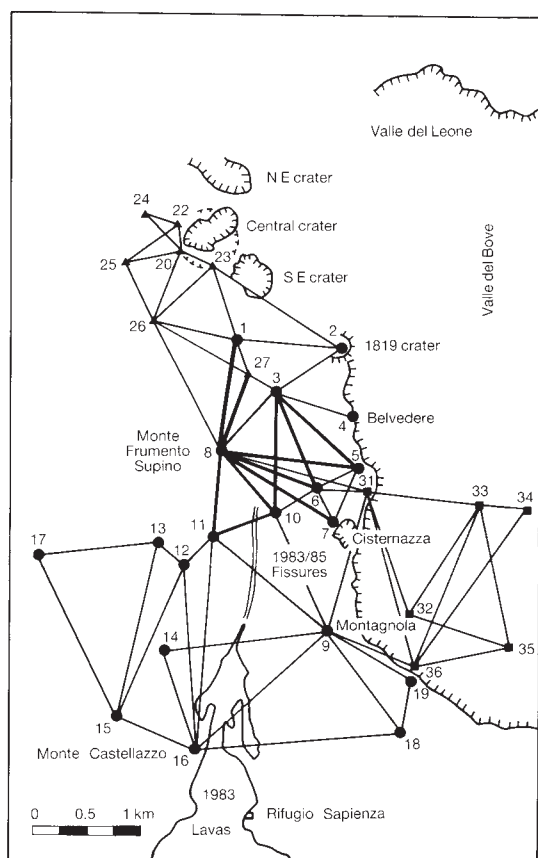


FIG. 1 The Etna geodetic network showing the principal lines measured, and the years during which the stations were established. $\blacktriangle$, Stations established during 1971; $\bullet$, stations established during 1981; $\blacksquare$, stations established during 1987. Cumulative length changes (1981–1988) along the bold lines are shown in Fig. 2.

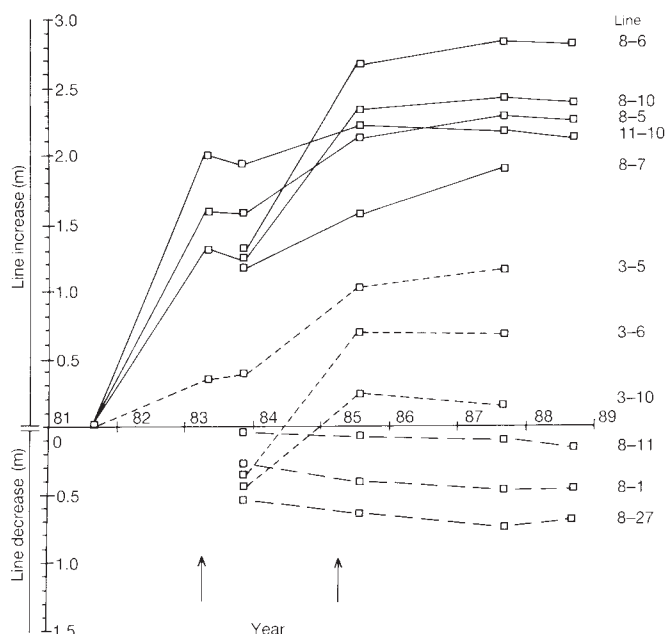


FIG. 2 Cumulative slope distance changes for selected lines (see Fig. 1) between 1981 and 1988. Continuous lines represent network lines, orientated roughly west-east, which cross the 1983 or 1985 fissures and feeder-dyke systems. Large dashes represent lines running north-south which parallel the fissure-dyke systems. Small dashes represent lines of intermediate orientation. The opening of the 1983 and 1985 fissures are marked by the arrows.

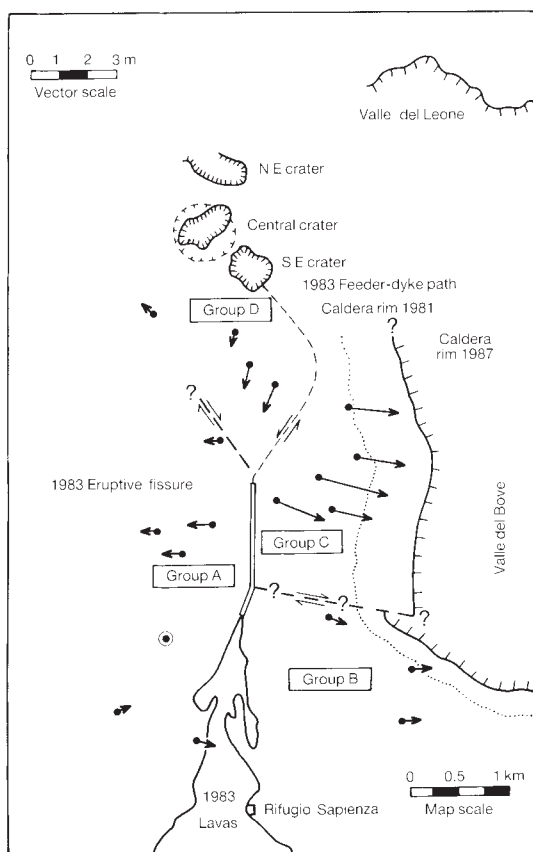


FIG. 3 Horizontal ground movements displayed in terms of cumulative coordinate changes (1981–1987) relative to Station 14. Coherent-group displacement vectors are explained within a framework of block movement attributed to brittle fracture and dilation associated with dyke emplacement. The sub-surface path of the 1983 eruption feeder dyke is marked by small dashes, and the large dashes indicate the postulated boundaries of discrete structural blocks.

of up to 30 cm despite there being no further dyke emplacement. One possible explanation of this behaviour is the continued eastward movement of part of the caldera rim due to gravity, although measurements recorded in July 1988 show that this movement has now effectively ceased. The period between July 1987 and July 1988 is characterized generally by movements that are an order of magnitude less than those generated during and immediately following dyke emplacement. In many cases, line-length changes are barely significant, reflecting the absence of intrusive activity in the southern 'rift' over the period.

Horizontal ground movements displayed in terms of cumulative coordinate changes between 1981 and 1987 indicate that the stations can be divided into four groups on the basis of coherent displacement behaviour. We summarize the group displacement characteristics in Table 1 and Fig. 3 shows displacement vectors relative to station 14, which we regard as being distant enough from both the fissures and the caldera rim to act as a relatively stable base station.

The simplest explanation for coherent group behaviour in this context is provided by a mechanism of block movement initiated by dyke emplacement. Dieterich and Decker⁶ and Pollard *et al.*⁷ have modelled theoretically the vertical and horizontal ground displacements produced by elastic deformation above a shallow dyke. In the case of the 1983 feeder dyke, however, elastic deformation was only a precursor to brittle fracture, resulting in the formation of a 1.4-m-deep graben above

the dyke⁵. At least during this eruption, dyke emplacement led to fracturing of the upper southern flank, with fissures extending downwards (if the dimensional data of ref. 5 are accepted) to a depth of 1 km. The feeder dykes associated with the 1983 and 1985 eruptions therefore represent essentially planar surfaces separating discrete crustal blocks of ~1-km thickness.

We can explain coordinate changes over the period 1981–1987 most satisfactorily using a model in which proximity to the Valle del Bove caldera results in a modification of the gravity-controlled stress regime operating within the upper levels of the volcano. We have previously suggested^{2,3} that reorientation of stresses adjacent to the 1-km-high cliff wall of the caldera favours the preferential emplacement of high-level dykes parallel to the western rim. If this is correct, the dilation associated with dyke emplacement in the area is accommodated largely by the eastward displacement of a block of caldera rim. Comparison between displacement vectors shown by Group C stations and those of Group B, demonstrates that the moving block is confined to the unbuttressed western wall. Group B station vectors are significantly smaller, indicating the relative stability of this part of the rim. We suggest that this is due to a buttressing effect caused by the alignment of the southern caldera rim, which is orientated normal, rather than parallel, to the 1983 and 1985 fissures. The disparity in the degree of displacement between the two groups of stations indicates the existence of a zone of dextral shear between the moving block to the north and the stable block to the south. To the west of the fissures, Group A stations display progressively decreasing displacement vectors with increasing distance from the fissures indicating that some dilation is accommodated by crustal shortening. This implies that the dyke intrusion here is a dynamic event, rather than one involving the passive entry of magma into incipient open fractures.

Based on the 2.8-m displacement of the western caldera rim block over six years, we believe that the potential for future, if not imminent, slope failure is increasing. At present, there are insufficient data to predict either the likelihood of such an event taking place, or the form a slope failure might take. Rotational slip along incipient fracture surfaces paralleling the caldera rim, or along the contacts of newly intruded dykes are likely mechanisms. A worst-case scenario, in which slope-failure would be initiated by rotational slip could involve magma entering a deep curving fracture parallel to the caldera rim. Rotational movement along such a surface has the potential to generate a catastrophic eruption such as that of the Mount St Helens in May 1980 in which a blast event was accompanied by dry avalanches, pyroclastic flows and lahars. Should the failure occur in winter or early spring, the contribution of snow melt will dramatically increase the run-out distances of debris flows.

Assuming maintenance of the stress regime now operating within the upper part of the southern 'rift zone', we predicted³ that further dyke emplacement would be likely to be parallel to the western caldera rim, as long as magma continued to rise to the south-east of the present summit. Each further intrusion event in the area increases not only the potential for slope failure, but also the risk of inundation of inhabited areas by the accompanying debris flows. □

Received 6 September; accepted 6 December 1989.

- McGuire, W. J. thesis Luton Coll. Higher Educn (1980).
- McGuire, W. J. & Pullen, A. D. *Geol. Soc. Lond. News* (abstr.) **16**, 42 (1987).
- McGuire, W. J. & Pullen, A. D. *J. Volcan. geotherm. Res.* **38**, 325–344 (1989).
- Wadge, G. *J. Volcan. geotherm. Res.* **1**, 237–263 (1976).
- Murray, J. B. & Pullen, A. D. *Bull. volcan.* **47**, 1145–1163 (1984).
- Dieterich, J. H. & Decker, R. W. *J. geophys. Res.* **80**, 4094–4102 (1975).
- Pollard, D. D., Delaney, P. T., Duffield, W. A., Endo, E. T. & Okamura, A. T. *Tectonophysics*. **94**, 541–584 (1983).

ACKNOWLEDGEMENTS. We thank Sokkisha UK, Imperial College and the Istituto di Vulcanologia (Catania) for the supply of surveying equipment. We acknowledge the support of NERC and The West London Institute for financing field work.

Book-lungs in a Lower Carboniferous scorpion

Andrew J. Jeram

Department of Geology, University of Manchester, Manchester, M13 9PL, UK

INDIRECT evidence indicates that scorpions, which first appeared in the middle Silurian, were originally aquatic organisms like their eurypterid relatives¹. Living scorpions have four pairs of book-lungs, each pair situated above a sternite on the ventral surface of the mesosoma (anterior abdomen) and each book-lung opening to the outside through a stigma which perforates the sternite. By contrast, most Palaeozoic scorpions had five abdominal plates, homologues of abdominal appendages, which were apparently sutured onto the body wall only along their anterior edges. It has been suggested that there were gills above the abdominal plates^{2,3} and that all scorpions with abdominal plates were aquatic and respired through gills^{2,4}. The only good example, however, of a Palaeozoic scorpion with gill-like structures preserved is the Lower Devonian *Waeringoscorpio*^{5,6}. Portions of book-lungs have now been discovered in two specimens of a fossil scorpion with abdominal plates from a Lower Carboniferous limestone in Scotland, providing the first direct evidence of book-lungs and also the earliest evidence of air-breathing in a Palaeozoic scorpion. Unlike Recent scorpions, the fossil lung lamellae have ribs of thicker cuticle along their posterior margins, supporting the homology of these structures with the *Limulus* book-gill^{3,7}. As Silurian and Devonian scorpions were aquatic^{1,2} the presence of book-lungs in a Carboniferous scorpion indicates that the transition from aquatic to terrestrial environments was achieved by the direct transformation of book-gills into book-lungs.

More than 20 specimens attributable to a single new scorpion

genus have been recovered from a Dinantian (Lower Carboniferous) limestone at East Kirkton Quarry, West Lothian, Scotland, along with an assemblage of terrestrial vertebrates, eurypterids, millipedes and plants⁸. The finely laminated limestone is interbedded with tuffs and is thought to have been deposited in a lagoonal environment which was influenced by contemporaneous hot springs and local volcanic activity⁹. Two scorpion specimens have parts of the book-lungs preserved. Both specimens are preserved as thin, brown, organic cuticle, which probably represents only the hyaline exocuticle layer of the original cuticle¹⁰. The preservation of the book-lung cuticle and some arthroal membrane is different from that of the rest of the integument, consisting of porous white residues which may indicate replacement by silica, although no data on their chemical composition are available at present. The smaller specimen (NMS Geol. 1985.4.67) has five pairs of lamellate structures above the abdominal plates and has been briefly discussed elsewhere¹¹, although very little detail of these book-lungs can be discerned. The larger individual (NMS Geol. 1987.7.136; Figs 1 and 2) was approximately 30 cm long in life, and portions of the posterior four pairs of book-lungs have been recognized. Unlike NMS Geol. 1985.4.67, the vertically arranged lamellae had largely decomposed before cementation of the sediment, leaving only thickened ribs of cuticle along the posterior end of each lamella preserved (Fig. 2a). Where small portions of the lamellae are present (Fig. 2c), they lie on the anterior side of the ribs and project into the body cavity. Each rib is fused at its dorsal and ventral ends to a flattened elliptical annulus of cuticle (Figs 1 and 2c). Together, the ribs and annulus formed a rigid grille at the posterior (atrial) end of the book-lung. The atrium consists of the space between the body wall and the underlying abdominal plate posterior to each book-lung.

The ribs of the fossil book-lungs are probably homologous to the marginal thickenings of the gill lamellae of the living xiphosuran *Limulus*¹². They bear conical tubercles on the atrial

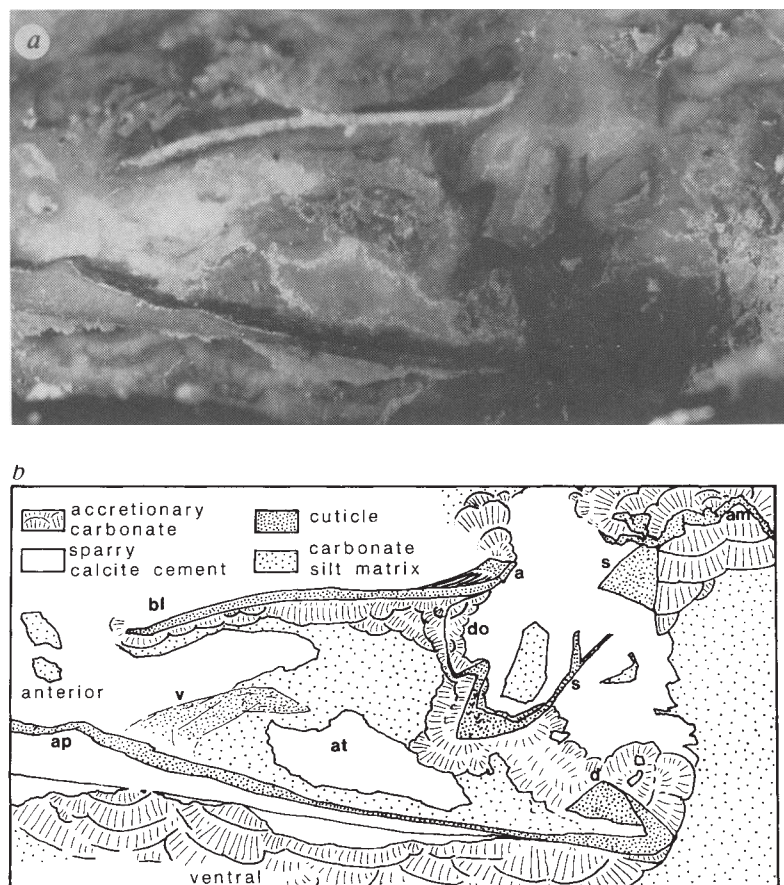


FIG. 1 Scottish Dinantian scorpion, specimen NMS Geol. 1987.7.136. Vertical section through the left lobe of abdominal plate 4, showing the rigid grille of the left book-lung (bl), true sternite (s), and abdominal plate (ap). The section has been acid-etched so that the cuticle stands up in relief. Note that compaction has distorted the integument. The ventral wall of the atrium (v) is no longer attached to the book-lung grille, and the true sternite (s) is broken into two fragments in this plane of section. do, Dorsal wall of atrium; ap, ventral wall of abdominal plate; at, atrium; d, doublure of abdominal plate; am, arthroal membrane; a, annulus. a, Detail of atrial area, photographed dry in incident light ($\times 11$). b, Simplified diagram showing same area as a ($\times 11$).

side (Fig. 2a,b) which may be the bases of reduced fringing hairs like those present in *Limulus*. The ventral surface of the atrium is formed by the dorsal wall of the abdominal plate, and the dorsal surface by a membranous cuticle which grades posteriorly into the true sternite (Fig. 1). The true sternite forms a broad crescent-shaped sclerite which lies above the posterior margin of the abdominal plate. It is bordered posteriorly by arthrodial membrane which connects it to the succeeding abdominal plate, and laterally by arthrodial membrane connected to the dorsal tergites. The abdominal plate is sutured onto the ventral body wall in the anteromedian area, although posterior to the book-lungs it is free. Air could therefore enter the atrium along the entire posterior margin of each plate and no stigmata are present. The rigid nature of the grille, and the projection of lamellae into the body cavity rather than into the chamber (atrium) above the abdominal plate, would prevent the free circulation of water across the lamellae, precluding the possibility that these structures functioned as book-gills rather than book-lungs.

This material indicates that the possession of abdominal plates is not by itself an indicator of an aquatic lifestyle, and that most late Palaeozoic scorpions could have been fully terrestrial.

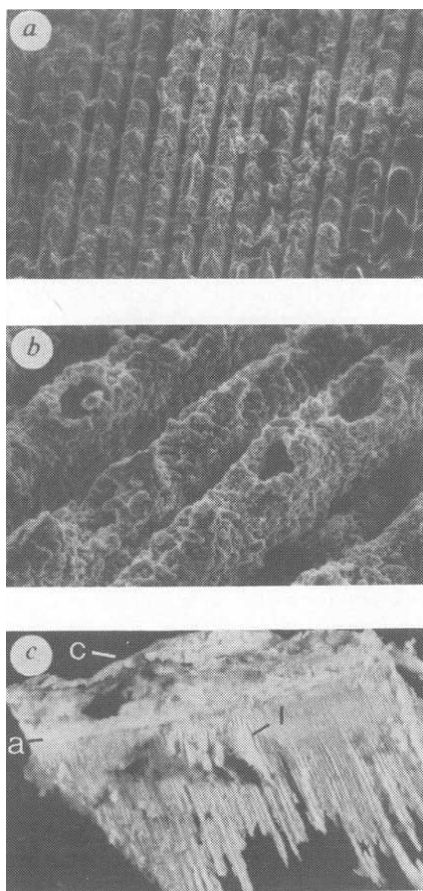


FIG. 2 Scottish Dinantian scorpion, specimen NMS Geol. 1987.7.136. Details of the anterior and posterior surfaces of fragments of the book-lung grille, etched free from the matrix with dilute HCl. a, Scanning electron micrograph of the posterior (atrial) side of a fragment of book-lung grille from the third pair of book-lungs showing the rigid bars of cuticle lined with the bases of short spines or fixed setae ($\times 415$). b, Scanning electron micrograph showing detail of same specimen as (a). Where the fringing hairs are broken away, the hollow sinus within each bar can be seen ($\times 1,450$). c, Anterior side of a fragment of the book-lung grille from the fifth pair of book-lungs, photographed dry in incident light. Small portions of some lung lamellae (l) are still preserved in this specimen. a, Annulus; c, thin cuticle of atrial wall ($\times 15$).

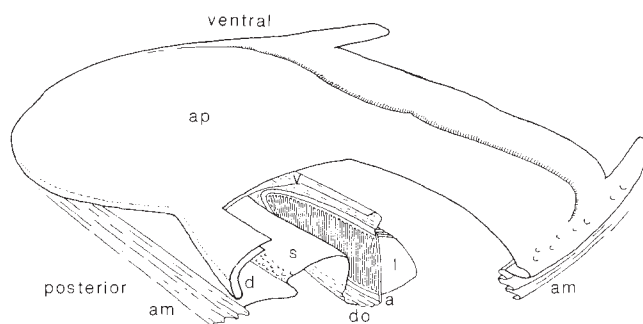


FIG. 3 Reconstruction of an abdominal plate, sternite and book-lung of Scottish Dinantian scorpion. Viewed from below, with abdominal plate cut away to reveal dorsal structures. Based on Royal Museum of Scotland specimens NMS Geol. 1987.7.36, NMS Geol. 1985.4.67 and NMS Geol. 1985.4.66. Labels are as in Figs 1 and 2.

Recent scorpions do not possess a rigid grille at the posterior ends of their book-lungs; instead, the distal ends of lamellae grade directly into the atrial wall dorsally and ventrally. The rigid grille present in the fossil material probably prevented lamellae collapsing together, thereby ensuring an adequate respiratory surface area.

The earliest terrestrial chelicerates date from the Lower and Middle Devonian¹³, at a time when other evidence indicates that scorpions were aquatic¹¹. Therefore, scorpions almost certainly came onto land independently of other chelicerates, and book-lungs must have evolved more than once. The only other record of book-lung preservation in a Palaeozoic arachnid is in Lower Devonian trigonotarbid from the Rhynie chert of Aberdeenshire¹⁴. These appear to be similar to those of modern spiders and do not add to our understanding of how the parallel evolution of book-lungs in two or more arachnid groups occurred.

Stigmata, and therefore presumably book-lungs, occur in *Palaeopisthacanthus* from the Upper Carboniferous of Illinois¹⁵. *Palaeopisthacanthus* is almost certainly not a descendant of the older East Kirkton scorpions, and may represent a separate origin for stigmata¹¹ or even book-lungs, perhaps in a group of scorpions more immediately ancestral to the modern forms. The presence of five pairs of book-lungs in the East Kirkton scorpions, and the similarity of the respiratory lamellae with those of *Limulus*, may be regarded as primitive characteristics, and provide the strongest evidence yet that arachnid book-lungs were derived directly from book-gills in their aquatic ancestors. □

Received 29 September; accepted 21 November 1989.

1. Rolfe, W. D. I. & Beckett, C. M. *Spec. Pap. Palaeont.* **32**, 27–37 (1984).
2. Kjellesvig-Waering, E. N. *Palaeontogr. Am.* **55**, 1–287 (1986).
3. Størmer, L. *Serkenbergiana lethaea* **57**, 87–183 (1976).
4. Briggs, D. E. G. *Nature* **326**, 645–646 (1987).
5. Størmer, L. *Serkenbergiana lethaea* **51**, 335–369 (1970).
6. Brauckmann, C. *Geol. Palaeont.* **21**, 73–85 (1987).
7. Lankester, E. R. Q. *J. micros. Sci.* **25**, 339–342 (1885).
8. Wood, S. P., Panchen, A. L. & Smithson, T. R. *Nature* **314**, 355–356 (1985).
9. Rolfe, W. D. I. *Earth Sci. Conserv.* **25**, 22–28 (1988).
10. Bartram, K. M., Jeram, A. J. & Selden, P. A. *J. geol. Soc. Lond.* **144**, 513–517 (1987).
11. Selden, P. A. & Jeram, A. J. *Trans. R. Soc. Edinb.* **80**, 303–310 (1989).
12. Snodgrass, R. E. *A Textbook of Arthropod Anatomy* (Cornell University Press, New York, 1952).
13. Shear, W. A. et al. *Am. Mus. Novit.* **2901**, 1–74 (1987).
14. Claridge, F. M. & Lyon, A. G. *Nature* **191**, 1190–1191 (1961).
15. Vogel, B. R. & Durden, C. J. *J. Palaeont.* **40**, 655–658 (1966).

ACKNOWLEDGEMENTS. I thank S. P. Wood and W. D. I. Rolfe for the opportunity to prepare material from the East Kirkton Quarry and P. A. Selden for helpful comments on the manuscript. This work was carried out whilst in receipt of an NERC Postdoctoral Research Fellowship.

Saccadic oscillations facilitate ocular perfusion from the avian pecten

John D. Pettigrew*, Josh Wallman†
& Christine F. Wildsoet‡

* Vision, Touch and Hearing Research Centre, University of Queensland, St Lucia 4067, Australia

† Department of Biology, City College of City University of New York, New York 10031, USA

‡ Department of Optometry, Queensland University of Technology, Brisbane 4001, Australia

THE evolution of the eye is constrained by two conflicting requirements—good vascular perfusion of the retina, and an optical path through the retina that is unobstructed by blood vessels. Birds are interesting in that they have higher metabolic rates and thicker retinas than mammals, but have no retinal blood vessels. Nutrients and oxygen must thus reach the neurons of the inner retina either from the choroid through 300 μm of metabolically very active retina, or from the pecten, a pleated vascular structure protruding from the head of the optic nerve into the vitreous chamber, and more than a centimetre away from some retinal neurons. Despite the diffusional distance involved, several lines of evidence indicate that the pecten is the primary source of nutrients for the inner retina: the presence of an oxygen gradient from pecten to retina¹, the large surface area produced by macroscopic folds^{2,3} and by microscopic infoldings of the luminal and external surfaces of the capillary endothelium⁴⁻⁶, extrusion of circulating fluorescein⁷, high content of carbonic anhydrase and alkaline phosphatase^{8,9}, and retinal impairments after pecten ablation¹. Another peculiarity of birds, their saccadic oscillations, occur with a large cyclotorsional component during every saccadic eye movement¹¹. In different species, saccades, which occur at intervals of 0.5–40 s, have up to 13 oscillations with frequencies of 15–30 Hz and amplitudes of about 10° (ref. 12). Therefore, as much as 12% of some birds' total viewing time may be subject to the image instability caused by the oscillations¹³. Using fluorescein angiography, we show here that during every saccade, the pecten acts as an agitator

which propels perfusate towards the central retina much more effectively than is observed during intersaccadic intervals.

Five species of birds were examined with indirect ophthalmoscopy after an injection of fluorescein into a wing vein: domestic chicken (*Gallus gallus*); tawny frogmouth, (*Podargus strigoides*); boobook owl (*Ninox novaeseelandiae*); bush thick-knee (*Burhinus grallarius*); and the laughing kookabura (*Dacelo gigas*). In all the birds examined, the fluorescent dye filled the pecten (Fig. 1a) within 4–10 s of injection and then leaked more slowly from the pecten into the vitreous chamber, thus confirming earlier reports^{7,14}. Our observations concern the effects of saccadic oscillations (Fig. 1b) on the outward movement of the dye from the pecten. A single saccade (Fig. 1b) propelled the dye, in a plume or series of wavelets, to distances that were an order of magnitude further from the pecten than had occurred in the entire previous intersaccadic interval. This was most easily seen in the thick-knee and the frogmouth, neither of which tended to blink during a saccade.

An experiment on a frogmouth is illustrated in Fig. 2. Six seconds after intravenous injection as a bolus, fluorescein filled the pecten—a visually dramatic event as the previously black pecten turned bright green and its folds became evident. Subsequently, the dye seemed to leak from the pecten into the vitreous chamber to form a slowly intensifying halo at the base of the pecten close to the retina. Leakage of dye from the pecten became noticeable at 40 s. More than 1 min after leakage was first observed, the outer edge of the dye had moved less than 0.5 mm from the margins of the pecten. A saccade at this point dispersed the halo and caused a plume of dye to be propelled in a superior and nasalward direction (Fig. 2). During the saccade, the pecten was observed to oscillate orthogonally to its long axis as might be expected if, as in chickens, oscillations are mainly cyclotorsional¹⁰. At the end of the saccade, the plume of dye extended more than 3 mm (the length of the pecten) and the edge of the dye at the lateral margins of the pecten had also moved outwards by 1–2 mm. Each subsequent saccade was accompanied by plumes and by stepwise extensions of the outermost edge of the dye. Similar plumes have been observed during fluorescein angiography of pigeon and rhea, but no association with eye movements was made¹⁴.

Although the infrequent pulsatile dispersion of fluorescein in the frogmouth makes the saccadic facilitation of perfusion dra-

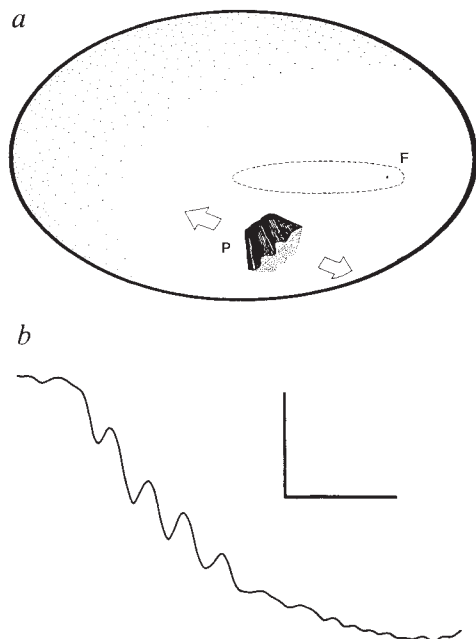


FIG. 1 a, Pecten oculi and relations in the left eye of the tawny frogmouth. The pecten is the folded pigmented structure (P) that projects into the eye from the optic nerve head. Its base is ~3 mm long and it is 3 mm high. In the frogmouth, there is a single binocular fovea in temporal retina (F). The fovea is located within a region of high ganglion cell density which extends towards the nasal retina (dotted ellipse). The arrows indicate the transverse movements of the pecten produced by the torsional component of the saccadic oscillations. The boobook owl had an arrangement very similar to that shown here, except that the pecten had an extra fold on its inferior limb. Other birds may have longer pectens with more folds and a second (monocular) fovea located in the central retina near the nasal end of the zone of high density. The thick-knee had a pecten with 11 folds, but the superior limb had the same position as shown here for the frogmouth pecten, which has three folds (that is, the extra eight folds are added to the inferior limb of the pecten to maintain a comparable relationship between the superior tip of the pecten and the retinal horizon). The thick-knee had a monocular fovea in a position corresponding to the nasal end of the ellipse of high ganglion cell density shown. The kookabura had an extra 17 pecten folds added to the inferior limb, and two foveas, a binocular fovea in the same relative position as the frogmouth's, and a monocular fovea like the thick-knee, with a position corresponding to the nasal end of the dotted region of increased density. b, An oscillatory saccade from the tawny frogmouth. Recording obtained with a sub-conjunctival search coil as described by Wallman and Pettigrew¹². Only the horizontal component is shown. The actual amplitude of the oscillations would be larger if the coil had been orthogonal to the cyclotorsional plane in which the oscillations occurred. Down is motion to the right. Horizontal scale, 250 ms; vertical scale, uncalibrated, ~5°. In frogmouths, the number of oscillations in a given saccade varies from five to ten and the mean intersaccadic interval is 40 s.

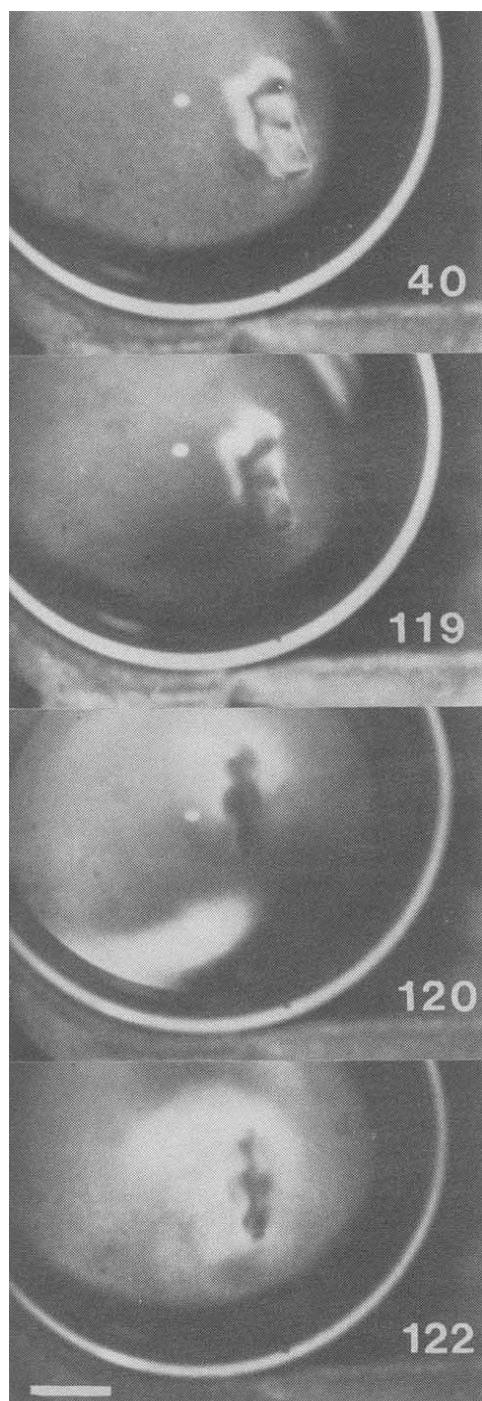


FIG. 2 Fundus of tawny frogmouth: sequence at various times (indicated in seconds) after intravenous injection of fluorescein. In the fundus at 40 s the pecten stands out as brightly fluorescent with a small amount of leaked dye accumulating as a halo around the pecten. Dye leakage is restricted to within 0.5 mm of the pecten. At 119 s just before the first saccade, the dye is still restricted to within 0.5 mm of the pecten. At 120 s, during the first saccade following the injection, a plume of dye has been propelled upward and nasally from the superior limb of the pecten. At 122 s, immediately after the saccade, the dye has been propelled 2–3 mm from the pecten. Note that the intersaccadic interval of the frogmouth was particularly long in this experimental situation, thus providing a clear view of perfusion in both situations, when saccades were either present or absent. In other situations and other birds, intersaccadic intervals were shorter, but the same phenomena were also observed. Mean intersaccadic intervals for the species were: chicken, 1.3 s; kookaburra, 8.8 s; tawny frogmouth, 40 s; boobook owl, 34 s; bush thick-knee, 16 s. Scale bar, 2 mm.

METHODS. Pupils were dilated by application, at least 1 h before examination, of proxymetacaine drops (Ophthalmic Allergan); followed by 1 drop min^{-1} of vecuronium bromide (4 mg ml^{-1} , Organon Teknika B.V.) for 20 min (as well as providing local anaesthesia, the proxymetacaine drops improved penetration of the curariform drugs). The procedure was effective in producing wide pupillary dilation for all species except the kookaburra and thick-knee, whose pupils were dilated with an injection of $50 \mu\text{l}$ of the vecuronium solution directly into the anterior chamber. The fundus in the region of the pecten was visualized by focusing an operating microscope (Wild Leitz M-650) on the inverted aerial image produced by a 30 D aspheric field lens clamped just in front of the eye. General anaesthetic agents were avoided because of their inhibitory effects on eye movements, particularly the saccadic oscillations. Local anaesthetic (4% lignocaine) was sprayed on the venipuncture site over the brachial vein. Sodium fluorescein (50 mg as a 100 mg ml^{-1} solution, Fluorescein Alcon) was injected as a bolus. Efflux of fluorescein from the pecten was monitored using a Wratten 47A excitation filter placed in the path of the illuminator of the operating microscope. Two beam splitters on the operating microscope enabled the sequence of changes in the fundus to be recorded on video tape and on 35 mm film.

rotation during the saccadic oscillations passes about a millimetre beyond the superior limb of the pecten (J. C. Letelier, personal communication).

The diverse shapes and dispositions of avian pectens have excited much unresolved speculation about their function^{15–18}. In the light of the present observations, the fan-shape and folds of most pectens take on new significance as potential propulsive surfaces, the arrangement of which must be coupled with the pattern of saccadic oscillations to produce the fluid motion we have observed. We propose that saccadic oscillations have co-evolved with pecten shape to optimize retinal perfusion. As a result, birds may have been able to exploit a range of ocular designs unavailable to mammals, such as the thicker, more complex avian retina and the achievement of panoramically high optical quality and visual acuity¹⁹. □

Received 11 October; accepted 21 November 1989.

1. Wingstrand, K. G. & Munk, O. *Biol. Skr. Dan. Vid. Selsk.* **14**, 1–64 (1965).
2. Braekveldt, C. R. *Ophthalmologica* **189**, 211–220 (1984).
3. Wood, C. *The Fundus Oculi of Birds* (Lakeside, Chicago, 1917).
4. Tucker, R. *Cell Tiss. Res.* **157**, 457–465 (1975).
5. Raviola, E. & Raviola, G. *Am. J. Anat.* **120**, 427–462 (1967).
6. Seaman, A. R. & Storm, H. *Expl Eye Res.* **2**, 163–172 (1973).
7. Abelsdorff, G. & Wessely, K. *Arch. Augenheilk.* **64**, 65–124 (1909).
8. Bawa, S. R. & YashRoy, R. C. *Expl Eye Res.* **13**, 92–97 (1972).
9. Amemiya, T. Graefe's *Arch. Clin. Exp. Ophthalmol.* **219**, 11–14 (1982).
10. Letelier, J. C. & Wallman, J. *Soc. Neurosci. Abstr.* **10**, 912 (1984).
11. Letelier, J. C., Evinger, C. & Wallman, J. *Soc. Neurosci. Abstr.* **13**, 172 (1987).
12. Wallman, J. & Pettigrew, J. D. *J. Neurosci.* **5**, 1418–1428 (1985).
13. Turkel, J. & Wallman, J. *Soc. Neurosci. Abstr.* **3**, 158 (1977).
14. Bellhorn, R. W. & Bellhorn, M. S. *Ophthalmic Res.* **7**, 1–7 (1975).
15. Barlow, H. B. & Ostwald, T. J. *Nature* **236**, 88–90 (1972).
16. Pettigrew, J. D. in *Animal Migration, Navigation and Homing* (eds Schmidt-Koenig, K. & Keeton, W.) 42–54 (Springer, Berlin, 1978).
17. Walls, G. L. *The Vertebrate Eye and its Adaptive Radiation*, 648–659 (Hafner, New York, 1967).
18. Duke-Elder, S. *The Eye in Evolution* Vol. 1, 362 (Mosby, St. Louis, 1958).
19. Raymond, L. *Vision Res.* **25**, 1477–1491 (1985).

ACKNOWLEDGEMENTS. We thank Laurie Hirst for advice on angiography and for providing the fluorescein, and Alan Cody, Rita Collins, Eddie Matejowski and Danny Thomas for technical assistance. We thank Peter Wilson and Julie Mason for the loan of birds. The work was supported by a Commonwealth Special Research Grant from the ARC to J.D.P., from the NSF to J.W. and grants from the NH&MRC and ARC to C.F.W. and J.D.P.

matic to observe, the fluctuations in gas exchange might not be so large because oxygen and carbon dioxide are smaller molecules than fluorescein and, because in most birds the pecten is larger, the eye smaller and the saccades more frequent. The advantage of using the frogmouth and thick-knee was that their large eyes and small pectens permitted visualization of the entire pecten and the surrounding retina in one field of view. We were therefore able to observe that dye movement created by oscillations was always directed upwards, roughly parallel to the long axis of the pecten and outward from its less-folded sweeping upper limb, a pattern we subsequently confirmed in the chicken. In other words, the combined action of pecten and oscillations is to direct the flow against gravity and towards the high-density regions of the retina. Consistent with this observation, quantitative studies of chickens show that the instantaneous axis of

A point mutation of the rhodopsin gene in one form of retinitis pigmentosa

Thaddeus P. Dryja, Terri L. McGee, Elias Reichel, Lauri B. Hahn, Glenn S. Cowley, David W. Yandell, Michael A. Sandberg & Eliot L. Berson*

Howe Laboratory of Ophthalmology and the Berman-Gund Laboratory for the Study of Retinal Degenerations, Harvard Medical School, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114, USA

* To whom correspondence should be addressed

THE gene for autosomal dominant retinitis pigmentosa in a large pedigree of Irish origin has recently been found to be linked to an anonymous polymorphic sequence, D3S47 (C17), from the long arm of chromosome 3 (refs 1, 2). As the gene coding for rhodopsin is also assigned to the long arm of chromosome 3 (refs 3, 4) and is expressed in rod photoreceptors that are affected early in this blinding disease⁵, we searched for a mutation of the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. We found a C→A transversion in codon 23 (corresponding to a proline→histidine substitution) in 17 of 148 unrelated patients and not in any of 102 unaffected individuals. This result, coupled with the fact that the proline normally present at position 23 is highly conserved among the opsins and related G-protein receptors⁶, indicates that this mutation could be the cause of one form of autosomal dominant retinitis pigmentosa.

The incidence of retinitis pigmentosa in the United States is estimated to be approximately 1:3,500 births^{7,8}. About 43% of cases in Maine are from families with an autosomal dominant mode of transmission, 20% are autosomal recessive, 8% are X-linked, 23% are isolated cases, and 6% are undetermined (for example, adopted)⁸. Genetic heterogeneity is thought to exist within each hereditary pattern. For example, linkage studies have revealed at least two distinct genetic loci on the X chromosome⁹. Hence, retinitis pigmentosa is not one disease, but a group of diseases caused by mutant alleles at various loci within the human genome. No gene defect has been described previously for any form of retinitis pigmentosa. Patients typically present with night blindness and loss of midperipheral visual field; as their condition progresses, they lose their far peripheral visual field and eventually central vision as well. Signs on fundus

examination in advanced stages include retinal vessel attenuation, intraretinal pigment in the peripheral fundus, and waxy pallor of the optic disc¹⁰. Patients have abnormal light-evoked electrical responses from the retina (as detected by electroretinograms (ERGs))¹¹, even in the early stages and in the absence of visible abnormalities on fundus examination¹²; relatives with normal ERGs do not go on to develop the disease¹³. There is a widespread loss of photoreceptors in the advanced stages^{14,15}.

We selected 20 unrelated patients with autosomal dominant retinitis pigmentosa from families with affected members who had delayed rod ERG responses and either normal or delayed cone ERG responses¹⁶. We synthesized pairs of 20-base oligomers, each pair designed to amplify one or two exons of the rhodopsin gene, using the polymerase chain reaction. With the amplified DNA fragments as templates, we sequenced the exons in leukocyte DNA from these patients. Five of the 20 patients were heterozygous for the same C→A transversion within codon 23 (AD133 in Fig. 1 is an example). We subsequently synthesized two oligomers (19-mers) corresponding to the normal and mutant sequences that include codon 23. We tested whether these sequences hybridized with the amplified DNA from each of a total of 148 unrelated patients (including the original 20 patients) with autosomal dominant retinitis pigmentosa, as well as from 102 normal individuals who were unrelated to the patients. Seventeen patients, including the original five who showed the mutation by sequence analysis, carried the C→A base change, whereas none of the normal individuals carried it ($\chi^2 = 12.57$; $P < 0.001$). This result effectively rules out the possibility that this nucleotide change represents a DNA polymorphism with no relationship to the disease. We also collected blood samples from four large kindreds with autosomal dominant retinitis pigmentosa and investigated the inheritance of alleles determined by a dinucleotide repeat polymorphism known to exist within the rhodopsin gene¹⁷. Only two of the families were informative for the dinucleotide polymorphism and only one (designated by number 5850) gave results consistent with cosegregation of the polymorphism with the disease. Hence, there are families with this disease who do not show cosegregation, confirming that there must be gene locus heterogeneity in this disease. In family 5850 (of British ancestry), we observed the C→A mutation in all affected members but not in any of the unaffected members (Fig. 2).

Codon 23 normally codes for a proline within the amino terminus of rhodopsin (Fig. 3). The precise function of this region is unknown, but the proline at this position is invariant

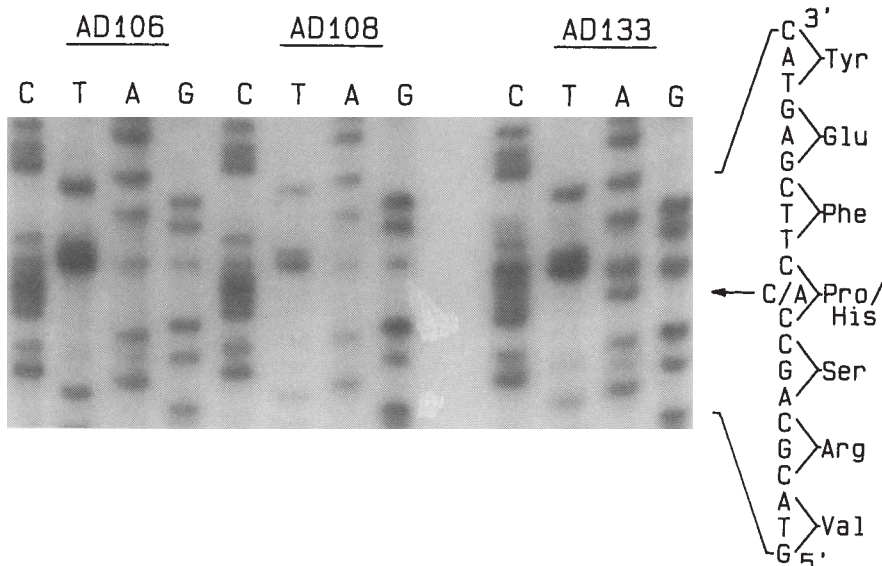


FIG. 1 Nucleotide sequence of codons 20 to 26 of the human rhodopsin gene in three patients with autosomal dominant retinitis pigmentosa. Patient AD133 is heterozygous for a C→A transversion within codon 23 (CCC to CAC). Patients AD106 and AD108 have the normal sequence in this region²¹.

METHODS. Using two oligonucleotide primers (5'-AGCTCAGGCCTTCGAGCAT-3' and 5'-GAGGCGTTGGATAACATTG-3') in the polymerase chain reaction, we amplified a 558-base pair sequence from 0.5 µg of leukocyte DNA from each patient. The amplified DNA fragment included exon 1 of the human rhodopsin gene²¹, and was used directly as a template for DNA sequencing²². In this case, the sequencing primer was the same as the 5' primer used for the polymerase chain reaction.

TABLE 1 Conservation of rhodopsin amino acid 23 among G-protein receptors

Amino acid	17	18	19	20	21	22	23	24	25	26	27	28	29
G-protein receptor													
Human rhodopsin	Thr	Gly	Val	Val	Arg	Ser	Pro	Phe	Glu	Tyr	Pro	Gln	Tyr
Bovine rhodopsin	Thr	Gly	Val	Val	Arg	Ser	Pro	Phe	Glu	Ala	Pro	Gln	Tyr
Human blue cone opsin	Ser	Ser	Val	—	—	Gly	Pro	Trp	Asp	Gly	Pro	Gln	Tyr
Human red cone opsin	Ser	Asn	Ser	Thr	Arg	Gly	Pro	Phe	Glu	Gly	Pro	Asn	Tyr
Human green cone opsin	Ser	Asn	Ser	Thr	Arg	Gly	Pro	Phe	Glu	Gly	Pro	Asn	Tyr
Drosophila R1-6 rhodopsin	Met	Ala	His	Leu	Ile	Ser	Pro	Tyr	Trp	Asn	Gln	Phe	Pro
Drosophila R8 rhodopsin	Met	Ala	His	Leu	Val	Asn	Pro	Tyr	Trp	Ser	Arg	Phe	Ala
Human β_2 adrenergic receptor	Ser	His	—	—	—	Ala	Pro	Asp	His	Asp	Val	Thr	Glu
Porcine M1 ACh receptor	Thr	Val	Leu	—	—	Ala	Pro	Gly	Lys	Gly	—	—	—
Porcine M2 ACh receptor	Leu	Ala	Leu	—	Thr	Ser	Pro	Tyr	Lys	Thr	—	—	—

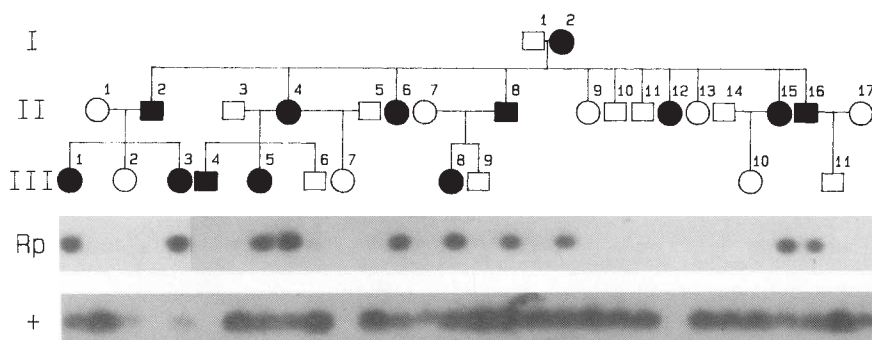
Codon numbers are for the normal human rhodopsin protein. The alignment of human rhodopsin with other opsins is according to ref. 6. Alignment of the human β_2 adrenergic receptor and the porcine M1 and M2 acetylcholine (ACh) receptors is according to ref. 20.

among the vertebrate and invertebrate opsins and in molecules sharing homology with the opsins, for example the beta-2 adrenergic receptor (Table 1)⁶. In view of the conservation of this amino acid, we expected the nucleotide change in codon 23 (substitution of histidine for nonpolar proline) to result in a dysfunctional or absent rhodopsin which would affect rod function. This prediction is consistent with ERG data from family 5850. Figure 4 illustrates normal ERG responses from an unaffected member (III-2, aged 28) and abnormal responses from two affected siblings and an affected aunt (III-3, aged 24; III-1, aged 29; and II-4, aged 52). The recordings in the left column show the response to flashes of dim blue light; this is a measure of rod function. The two affected siblings have a

markedly reduced response of the rods. The middle column shows the response to single flashes of white light which elicit a response from both rods and cones. Affected siblings have a splitting of the response into an early, cone-dominated peak and a delayed, rod-dominated peak of reduced amplitude (oblique arrows). The splitting occurs as a result of a relatively normal cone response time but a delayed rod response time, in contrast to the response of the 28-year-old normal member, for whom the two peaks cannot be distinguished. The right column shows the ERG response to flickering (30 Hz) white light; this is a measure of cone function, as only cones can respond to light flashes of this frequency. In comparison to responses of the normal member, the 24-year-old affected sibling

FIG. 2 Pedigree number 5850 with autosomal dominant retinitis pigmentosa. Solid symbols designate affected patients. The autoradiographs below the pedigree show the hybridization of each of two oligomers with an amplified DNA fragment, including exon 1 of the rhodopsin gene. The oligomer 'Rp' has the mutant sequence within codon 23 (5'-ACGCAGCCACTTCGAGTAC-3'). Oligomer '+' has the corresponding normal sequence (5'-ACGCAGCCCTTCGAGTAC-3'). Only the affected patients show intense hybridization with oligomer 'Rp'. There are blanks in the autoradiographs for individuals II-2, II-12, III-4 and III-7 because we did not have samples of their leukocyte DNA.

METHODS. Leukocyte DNA was amplified as described in the legend to Fig. 1. Amplified DNA fragments were purified by electrophoresis in a 2% agarose gel, then transferred to nylon membranes. Oligomers were labelled with ³²P



using polynucleotide kinase. Procedures for hybridization and washing were modifications of those previously described^{23,24}.

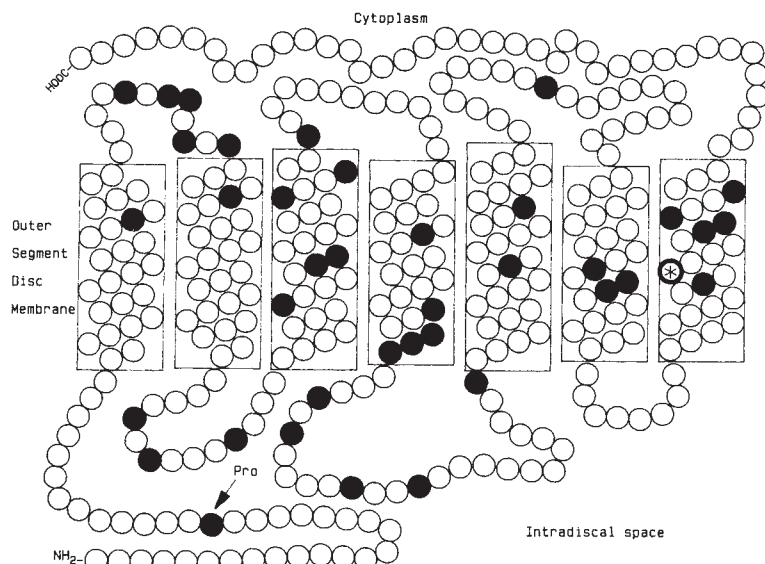


FIG. 3 Schematic representation of opsin in the rod outer segment disc membrane. Arrow designates the predicted site of mutation near the amino terminus of this protein⁶. Site of attachment of opsin to retinal to form rhodopsin(*), located in the seventh transmembrane segment, is separate from the site of this mutation. Shaded circles indicate amino acids that are invariant among vertebrate opsins.

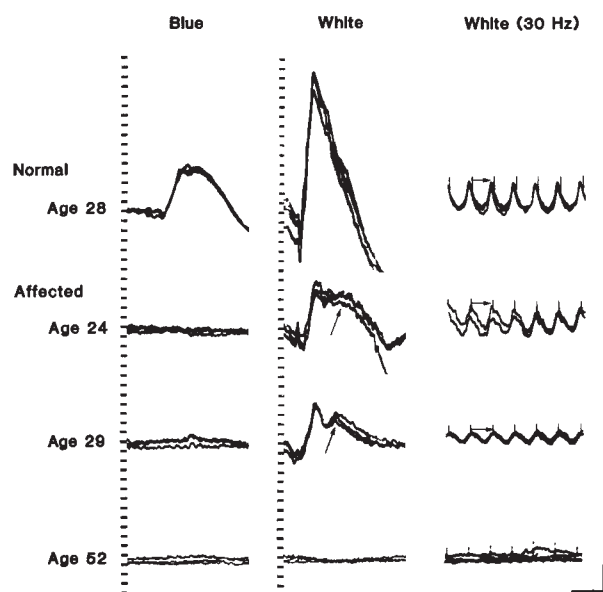


FIG. 4 Full-field ERGs from an unaffected patient (aged 28), her two affected siblings (ages 24 and 29 years), and an affected aunt (aged 52) in the family 5850 with autosomal dominant retinitis pigmentosa. Illustrated are rod-isolated responses to single flashes of blue light (left column), mixed cone-rod responses to single flashes of white light (middle column), and cone-isolated responses to 30 Hz white flickering light (right column), using techniques described previously^{18,25}. Stimulus onset, vertical hatched lines (left and middle columns) and vertical lines (right column). Two or three consecutive sweeps are superimposed. Cornea positivity gives an upward deflection. Oblique arrows in the middle column designate delayed rod-dominated peaks. Horizontal arrows in the right column designate cone response times (that is, the time interval between stimulus flash and corresponding cornea-positive response peak). Under these test conditions, normal amplitudes are $\geq 100 \mu\text{V}$ for single flashes of blue light, $\geq 350 \mu\text{V}$ for single flashes of white light, and $\geq 50 \mu\text{V}$ for 30 Hz white flicker; normal cone response times are $\leq 32 \text{ ms}$. Calibration symbol (lower right) designates 50 ms horizontally and 100 μV vertically.

demonstrates a normal amplitude and response time, whereas the 29-year-old affected sibling, with more advanced disease, has a slightly reduced amplitude and borderline delayed response time (horizontal arrows). The ERG results from the two affected siblings are consistent with the early, predominant involvement of rods in this form of autosomal dominant retinitis pigmentosa, as might be expected from a defect in rod-specific opsin. Late in this disease, there is loss of cone and rod function, illustrated by the profoundly reduced responses in all three columns from the 52-year-old affected aunt.

These ERGs correspond to those described for a family of British ancestry with autosomal dominant retinitis pigmentosa^{16,18}. One member of that family was among the first five patients who showed the mutation by sequence analysis (Fig. 1; AD133); most of our patients who carried the mutation were of British ancestry. On the other hand, patients of British ancestry from three other families with autosomal dominant retinitis pigmentosa¹⁹ did not show this mutation.

This mutation, which we designate 'rhodopsin 23, Pro $\rightarrow$ His,' appears to be the cause of one form of autosomal dominant retinitis pigmentosa. The mechanism by which this mutation, in a gene thought to be expressed exclusively in rods, leads to widespread degeneration of both rod and cone photoreceptors remains unclear. This discovery should encourage the evaluation of other genes coding for photoreceptor proteins for pathogenic mutations in different forms of retinitis pigmentosa. $\square$

- Nathans, J. *et al. Science* **232**, 203–210 (1986).
- Sparkes, R. S. *et al. Invest. Ophthalmol. Vis. Sci.* **27**, 1170–1172 (1986).
- Mandelbaum, J. *Arch. Ophthalmol.* **26**, 203–239 (1941).
- Applebury, M. L. & Hargrave, P. A. *Vision Res.* **26**, 1881–1895 (1986).
- Boughman, J. A., Conneally, P. M. & Nance, W. E. *Am. J. Hum. Genet.* **32**, 223–235 (1980).
- Bunker, C. H. *et al. Am. J. Ophthalmol.* **97**, 357–365 (1984).
- Chen, J. *et al. Am. J. Hum. Genet.* **45**, 401–411 (1989).
- Duke-Elder, S. & Dobrie, J. H. *System of Ophthalmology, X: Diseases of the Retina* (ed. Duke-Elder, S.) 574–622 (Mosby, St Louis, 1967).
- Karpe, G. *Acta Ophthalmol. (Suppl.)* **24**, 84 (1945).
- Goodman, G. & Gunkel, R. D. *Am. J. Ophthalmol.* **46**, 142–178 (1958).
- Berson, E. L. *Trans. Pa. Acad. Ophthalmol. Otolaryngol.* **26**, 109–113 (1973).
- Verhoeff, F. H. *Arch. Ophthalmol.* **5**, 392–407 (1931).
- Cogan, D. G. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* **54**, 629–661 (1949–1950).
- Berson, E. L. *Trans. Am. Acad. Ophthalmol. Otolaryngol.* **81**, 659–666 (1976).
- Weber, J. L. & May, P. E. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
- Berson, E. L., Gouras, P. & Gunkel, R. D. *Arch. Ophthalmol.* **80**, 58–67 (1968).
- Berson, E. L. & Simonoff, E. A. *Arch. Ophthalmol.* **97**, 1286–1291 (1979).
- Kobilka, B. K. *et al. Nature* **329**, 75–79 (1987).
- Nathans, J. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4851–4855 (1984).
- Yandell, D. W. & Dryja, T. P. *Cold Spring Harbor Symposium Series: Cancer Cells 7—Molecular Diagnostics of Human Cancer* (eds Furth, M. & Greaves, M.) 223–227 (Cold Spring Harbor Press, New York, 1989).
- Verlaan-de Vries, M. *et al. Gene* **50**, 313–320 (1986).
- Farr, C. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **85**, 1629–1633 (1988).
- Reichel, E., Bruce, A. M., Sandberg, M. A. & Berson, E. L. *Am. J. Ophthalmol.* **106**, 540–547 (1989).

ACKNOWLEDGEMENTS. We thank Amanda Bruce, Michelle Martinez, Jennifer Morrow, Basil Pawlyk, Peter Ringens, Charles Soparkar, and Carol Weigel for their assistance. This work was supported by grants from the Retinitis Pigmentosa Foundation Fighting Blindness, the George Gund Foundation, the National Eye Institute, awards from the Alcon Research Institute, and gifts to the Berman-Gund Laboratory and the Taylor Smith Laboratory at the Massachusetts Eye and Ear Infirmary. T.P.D. is a Research to Prevent Blindness—Dolly Green Scholar.

Possible role for salivary gland protein in taste reception indicated by homology to lipophilic-ligand carrier proteins

Hartwig Schmale, Heidi Holtgreve-Grez & Heidje Christiansen

Institut für Zellbiochemie und klinische Neurobiologie, Universität Hamburg, Martinistrasse 52, D-2000 Hamburg 20, FRG

SENSORY transduction in taste and olfaction, the principal chemical senses, seems to be mediated by membrane-associated proteins on the apical surfaces of the respective receptor cells^{1,2}. The recent isolation and characterization of soluble 'odorant-binding proteins' secreted from the nasal glands of rat³, cow⁴ and frog⁵, led to the hypothesis that these proteins function as necessary cofactors in olfactory transduction by concentrating and delivering odorants to the receptors⁶. The primary reception of taste stimuli occurs in specialized neuroepithelial receptor cells bundled in taste buds that are clustered in various types of papillae in the lingual epithelium of the tongue⁷. Small tubulo-alveolar salivary glands, the von Ebner's glands, are located beneath the circumvallate and the foliate papillae. Their ducts open exclusively into the trough at the base of the papillae^{8,9}. Taste buds located in the medial and lateral walls of the papillae open with their taste pores into the trough and consequently are in direct contact with the secretions of von Ebner's glands. Here we report the molecular cloning and characterization of a protein of relative molecular mass 18,000 that is highly expressed in von Ebner's glands. Like the odorant-binding proteins, this protein shows similarity to members of a protein superfamily of hydrophobic molecule transporters, indicating that pre-receptor events could also be necessary for the concentration and delivery of sapid molecules in the gustatory system, and emphasizing the close relationship of taste and olfaction.

Cell-free translation of poly(A)⁺ RNA isolated from adult male rat von Ebner's glands, including the circumvallate papilla, yielded an abundant protein product with a relative molecular mass of 20,000 ($M_r = 20\text{K}$; Fig. 1, lane 2). About 40% of the

Received 24 October; accepted 5 December 1989.

1. McWilliam, P. *et al. Genomics* **5**, 619–622 (1989).

2. Donis-Keller, H. *et al. Cell* **51**, 319–337 (1987).

clones isolated from a complementary DNA library prepared from von Ebner's gland poly(A)⁺ RNA contained inserts of 800 base pairs (bp) with identical restriction patterns. Sequence analysis of both strands of one of the clones, pta-1, revealed a single, long, open reading frame potentially encoding a protein of 177 amino acids with a calculated M_r of 19,726 (Fig. 2a). The initiator methionine of this protein is followed by a stretch of ~20 (mostly hydrophobic) amino acids characteristic of the signal peptide found in secretory proteins. The sequence contains no consensus sites for N-linked glycosylation.

Cell-free translation in the reticulocyte lysate and expression of the protein in frog oocytes after microinjection of von Ebner's gland poly(A)⁺ RNA further confirmed the identity of the cloned cDNA encoding a secretory non-glycosylated protein of the predicted size. *In vitro* SP6 polymerase transcription of clone pta-1 followed by cell-free translation of the synthesized sense RNA gave a single protein species of M_r 20K, identical in size

to the main product of von Ebner's gland poly(A)⁺ RNA translation (Fig. 1, lanes 2 and 4). Addition of dog pancreas microsomal membranes reduced the M_r of the pre-protein by ~2K because of cleavage of the signal peptide (Fig. 1, lanes 3 and 5). Also, a protein corresponding in size to the 18K processed form was present as a main secretory product in the incubation medium of *Xenopus laevis* oocytes primed with von Ebner's gland poly(A)⁺ RNA and in a protein extract prepared from male rat von Ebner's glands (Fig. 1, lanes 7 and 8).

RNA blots using a 200-bp fragment of clone pta-1 as a probe revealed an abundant 800-nucleotide RNA species in von Ebner's gland total RNA, indicating that clone pta-1 represents a full-length cDNA clone (Fig. 3a, lane 3). RNA coding for the von Ebner's gland protein (VEG protein) was detectable by

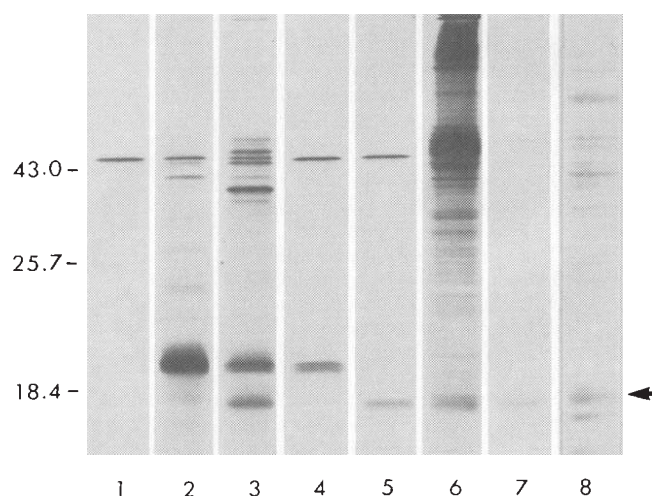


FIG. 1 Gel analysis of cell-free and oocyte translations of mRNA coding for the VEG protein. Lanes 1–5, [³⁵S]-methionine-labelled products of cell-free translations in the reticulocyte lysate; lanes 6 and 7, [³⁵S]-methionine-labelled proteins synthesized in frog oocytes after RNA microinjection. Lane 1, no mRNA added; lanes 2 and 3, translation of VEG poly(A)⁺ RNA without and with microsomal membranes, respectively; lanes 4 and 5, translation of RNA derived from *in vitro* transcription of clone pta-1 in the absence and presence of microsomal membranes, respectively; lane 6, proteins in oocyte homogenate after microinjection of VEG poly(A)⁺ RNA; lane 7, incubation medium of the microinjected oocytes; lane 8, Coomassie blue-stained soluble protein extract of rat von Ebner's glands. The arrow indicates the position of the processed 18K form of the VEG protein.

METHODS. For *in vitro* transcription, the insert of clone pta-1 was subcloned into plasmid pGEM4Z (Promega) after deleting the 5' dG·dC tail, and capped RNA was synthesized with SP6 polymerase. Cell-free translation in the reticulocyte lysate system (NEN) in the absence and presence of dog pancreas microsomal membranes as well as microinjection of *X. laevis* oocytes and incubation for 24 h in Barth's medium were as previously described¹⁸. Soluble proteins of von Ebner's glands were prepared by disrupting the rat tissue in an Ultraturax homogenizer in 20 mM Tris-HCl buffer, pH 7.5, and centrifuging at 5,000g for 10 min followed by 100,000g for 60 min to obtain clear supernatants for gel analysis. Protein products were characterized by 15% SDS-PAGE followed by staining with Coomassie blue and autoradiography; radioactive protein size markers ($M_r \times 10^{-3}$) were purchased from BRL.

FIG. 2 a, Sequence of the VEG protein cDNA and the deduced amino-acid sequence. Numbers of the nucleotides are given on the right of the individual lines. b, Alignment of the amino-acid sequence (single-letter code) of the VEG protein with the sequences of members of the lipophilic-molecule carrier protein family. Alignments were made using the FASTP algorithm¹⁹ as modified in the DNASTAR program Aalign. Ovine β -lactoglobulin (OBLG) was the protein with greatest similarity to VEG protein in the NBRF library (30% identity in a 183-amino-acid overlap). Similar scores were obtained when comparing the gustatory VEG protein with the olfactory proteins: the frog protein of Bowman's glands (BG 22% identity in a 178-amino-acid overlap) and the rat olfactory binding protein (OBPr 17% identity in a 175-amino-acid overlap). Sets with identical or conservative amino acids in at least three sequences are enclosed by solid lines.

METHODS. Double-stranded cDNA prepared by the RNaseH-DNA polymerase I method²⁰ using poly(A)⁺ RNA of male rat von Ebner's glands was inserted

a

```

GTCTCCTCAGTCTGCCAAGTGATCTGTCTTACCTGTGAGCAGACTCTGGAG ATG AAG GCC CTG CTC CTG ACC TTC GGC CTC AGC CTC CTG 94
Met Lys Ala Leu Leu Leu Thr Phe Gly Leu Ser Leu Leu

GCT GCC CTG CAG GCC CAG GCC TTC CCC ACC ACG GAA GAG AAT CAG GAT GTG TCA GGA ACG TGG TAT TTG AAA GCT GCA GCA 175
Ala Ala Leu Gln Ala Gln Ala Phe Pro Thr Thr Glu Glu Asn Gln Asp Val Ser Gly Thr Trp Tyr Leu Lys Ala Ala Ala

TGG GAC AAG GAG ATT CCT GAC AAG AAG TTT GGA TCT GTG TCT GTG ACT CCC ATG AAA ATG AAG ACC CTG GAA GGG GGC AAC 256
Trp Asp Lys Glu Ile Pro Asp Lys Lys Phe Gly Ser Val Ser Val Thr Pro Met Lys Ile Leu Thr Asp Leu Glu Gly Gly Asn

CTG CAA GTG AAG TTC ACT GTC CTG ATT GCA GGA CGA TGC AAG GAG ATG AGC ACT GTC CTA GAG AAG ACA GAT GAA CCT GCC 337
Leu Gln Val Lys Phe Thr Val Leu Ile Ala Gly Arg Cys Lys Glu Met Ser Thr Val Leu Glu Lys Thr Asp Glu Pro Ala

AAA TAC ACA GCC TAT AGT GGC AAA CAG GTA CTC TAC ATC ATA CCA TCT TCA GTG GAG GAC CAC TAC ATC TTC TAC TAT GAG 418
Phe Tyr Thr Ala Tyr Ser Gly Lys Gln Val Leu Thr Ile Thr Ser Ser Val Glu Asp His Tyr Ile Phe Tyr Tyr Glu

GGC AAA ATA CAC AGA CAT CAT TTC CAA ATT GCA AAA CTC GTG GGC AGA GAC CCT GAG ATC AAC CAG GAG GCC CTG GAA GAT 499
Gly Lys Ile His Arg His His Phe Gln Ile Ala Lys Leu Val Gly Arg Asp Pro Glu Ile Asn Gln Glu Ala Leu Glu Asp

TTT CAG AGT GTT GTA AGA GCT GGA GGC CTC AAC CCA GAC AAC ATC TTC ATC CCC AAG CAG AGT GAA ACA TGC CCT CTA GGA 580
Phe Gln Ser Val Val Arg Ala Gly Gly Leu Asn Pro Asp Asn Ile Phe Ile Pro Lys Gln Ser Glu Thr Gys Pro Leu Gly

AGC AAT TAG GGAAGTGATGCTCATACAGGCGCTGCTCAGCATTGCTCTACTGTCAGATCCACTGAGCTGGGACACAGAAGCAGCTCTTCACTGTC 684
Ser Asn Ter

CCTTCTGGCTCTGCCAACCTGCGCTCTGCTCCTCTTCTCTCTATAAAGAAGCTTCATCATC(A)n 753

```

b

```

VEG  MKALLLHLSLAALDAQAF-PT--TENVNDFHFWLKAANDKEIPDKKFGSVTPMKCTLEECNLQVFKFTVLIAGRCKEMSTLEK 88
OBLG  MKCLLLALILNACGVNHLIVTQTMKGLDQVNTWYSLANASDILSDAQSARFVYVVEKXPTPEHLEILLOKWNCEGCAKKILAEKTI
BG  MKLILHNLVLFPLQDADLPVPMKGLBNK-VITWVYHIAAENCKQFQIMKSDNAP-APNIVSLNNIMKSS-TSFQTEKGQGDVMEVNTI
OBPr  MVKFLNLLIALGVSGRHHENLDISPS--VNGQRTLVIVDNVEKVA-EGGSLRAYFQHEGDECEELKILFNVKLDSQDTHTVGQK

VEG  DEPAKK--TAYSGKQVLYLTPSSVEDHLYFYF-EGKILHRHH-FQIALVGRDPEINQALDEDPQSVVRAGGLNPDNIFPKQSET-PELGSN 177
OBLG  KIPAVFKINALNENKVL-VLQTDYKK-YLLFCMENSAEPOSQACQCLV-RTPEVDNEALEKFDKALKALPMHILRLAFNPQDDEGQCHV
BG  VEGHKKWKMOQDSETIEVATDYIA-FLMEFTT-KIQMGAEVCVTVKLFGKDKLIPEDMLKHPEDHIEKVLGKKEQ-VLRPHKATQVPK
OBPr  HDGRNLTDTYSGRNYFHLVKKTDIDIFPHNVNVDSEGRQ-CDLVA--GKRFDNKAQKQSLRKLAEEYNIPIENTQHLVPTDITLQD

```

into the plasmid vector pUC9 by dG·dC-tailing. The most abundant cDNA clones with identical restriction patterns were supposed to encode the VEG protein. Appropriate restriction fragments were subcloned into M13mp18 and M13mp19 vectors and DNA sequencing was carried out on both strands by the dideoxy chain-termination method²¹.

northern-blot analysis only in von Ebner's gland extracts, but not in other salivary glands, such as the parotid and mandibular glands or in liver and brain (Fig. 3a, lanes 1, 2 and 4). DNA blot analysis indicated that the VEG protein probably belongs to a small family of related proteins. The digestion of rat genomic DNA with various restriction enzymes yielded at least two different fragments each hybridizing with a 200-bp fragment of clone pta-1 (Fig. 3b).

The tissue-specific expression of the RNA coding for the VEG protein was further demonstrated by *in situ* hybridization with cross-sections of rat tongue. The messenger RNA was restricted to the acinar cells throughout the von Ebner's lingual salivary glands both at the levels of the circumvallate and of the foliate papillae (Fig. 4, a, b). Cells of the secretory ducts of the von Ebner's glands and of the mucus parts of the lingual glands did not give positive signals (Fig. 4c, d).

Comparison of the predicted amino-acid sequence of the VEG protein with all sequences in the NBRF-PIR protein data base (release no. 20) revealed similarities to a group of soluble 20K proteins that all belonged to a protein superfamily of lipophilic-molecule carriers¹⁰. Members of this family bind and transport small hydrophobic molecules such as retinoids, steroids, bilins and lipids, and have a particular three-dimensional structure consisting of an eight-stranded β -barrel and an α -helix¹¹. Com-

puter alignment of the amino-acid sequences indicated significant homologies between, for example, β -lactoglobulin, α_2 -microglobulin, retinol-binding protein and the olfactory-specific proteins BG (for Bowman's gland; ref. 5) and OBP (for olfactory binding protein; ref. 3) belonging to the same protein family (Fig. 2b). The VEG protein shares highly conserved motifs, such as Gly-X-Trp-Tyr (X is an unspecified amino acid), found in almost all members of the family, but varies considerably in those regions that are believed to be involved in determining ligand specificity.

In addition to the similarities between the predicted primary sequences of the olfactory binding proteins and the taste-associated VEG protein, the proteins are expressed in anatomically comparable sites in both systems in the respective epithelia: Whereas Bowman's glands produce the mucus that covers olfactory neurons, von Ebner's glands produce saliva that fills the furrows of the taste papillae.

These homologies indicate that the VEG protein of the gustatory system has a physiological role similar to that of OBP and BG proteins in olfaction. It has been postulated that the olfactory-specific proteins have a general function in concentrating and delivering airborne odorants to the olfactory receptors¹². In particular, they could be involved in the transport of highly lipophilic odorants across the hydrophilic mucus that covers

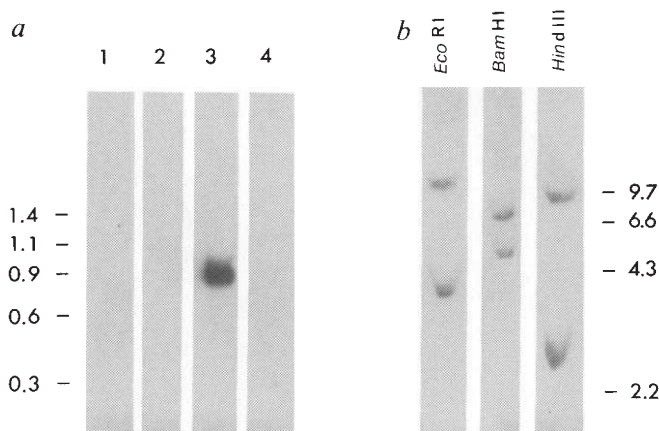
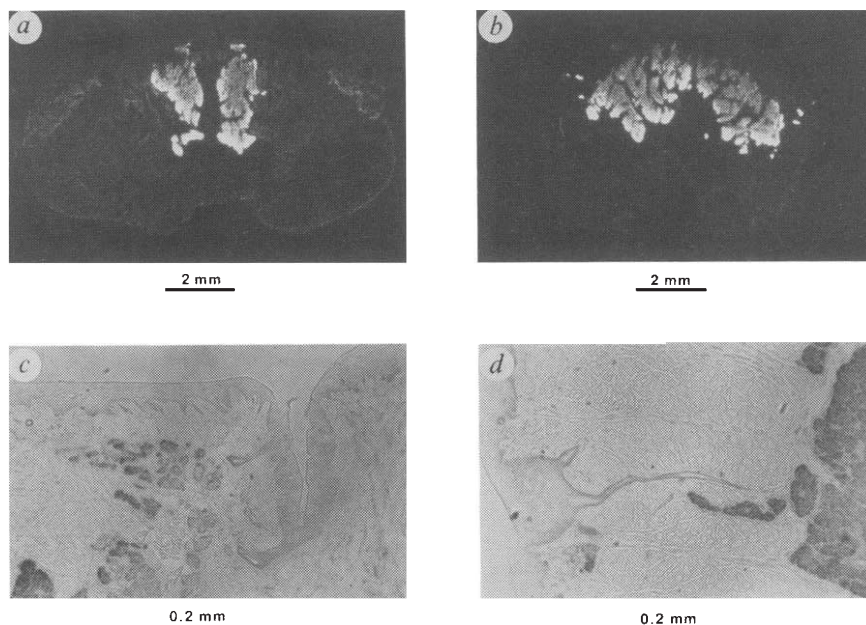


FIG. 3 RNA blot hybridization analysis of RNA from different tissues with a VEG protein cDNA probe. Each lane contained 15 μ g glyoxal-treated total RNA from the following rat tissues: lane 1, total brain; lane 2, parotid and mandibular gland; lane 3, von Ebner's gland; lane 4, liver. Size markers (kilobases) were derived from glyoxal-treated Φ X174 *Hae*III fragments (BRL). b, DNA blot hybridization analysis of rat genomic DNA after digestion with various restriction enzymes. Each lane was loaded with 5 μ g DNA restricted with the indicated enzymes. Molecular sizes were determined from λ *Hind*III fragments (BRL).

METHODS. After separation of glyoxylated RNAs on a 1.5% agarose gel and restricted genomic DNAs on a 0.8% agarose gel, nucleic acids were transferred to GeneScreen-plus membranes (NEN). After hybridization at 42 °C in 50% formamide and, 4 $\times$ SSPE where 1 $\times$ SSPE is: 10 mM sodium phosphate, pH 7.0, 1 mM EDTA, 180 mM NaCl with a 200-bp fragment of clone pta-1 labelled with [³²P]dCTP by random priming²², the blots were washed twice for 30 min at 22 °C followed by 15 min at 65 °C in 0.2 $\times$ SSC medium, 0.1% SDS, and analysed by autoradiography.

FIG. 4 *In situ* hybridization of VEG protein mRNA in cross-sections through rat tongue. a and b, Negative prints of X-ray autoradiograms of hybridized sections from the level of the single circumvallate papilla on the dorsum (a) and the foliate papillae on the sides of the tongue (b), respectively, showing heavy labelling of the entire serous von Ebner's gland extending ~3–4 mm on all sides of the papillae. c and d, Bright-field photomicrographs of sections through the circumvallate papilla (c) and a foliate papilla (d). Silver grains are present over secretory cells of von Ebner's gland close to the gutters of the papillae. Cells lining the ducts opening into the furrows and the gustatory epithelium bearing the taste buds do not show positive hybridization.

METHODS. Tongues of adult rats were dissected, embedded into Tissue-Tek (Miles), cut at 10 μ m in a cryostat and transferred to microscopic slides. Radiolabelled antisense RNA probes were generated by transcription using SP6 polymerase in the presence of [α -³⁵S]CTP from clone pta-2, a subcloned 200-bp *Pst*I–*Rsa*I fragment of clone pta-1. After overnight incubation at 50 °C in hybridization solution containing 50% formamide and 4 ng labelled antisense RNA per slide, the sections were treated with RNase A followed by washing in 2 $\times$ SSC medium at 50 °C (ref. 23). Sections were exposed for two days to X-ray film, dipped in NTB-3 photographic emulsion (Kodak) and developed after 3 days. Controls for the specificity of *in situ* hybridization (data not shown) included the use of radiolabelled sense RNA of clone pta-2 and of antisense



RNA for the unrelated vasopressin precursor. Neither of the probes gave positive hybridization signals on tongue sections containing von Ebner's glands.

the sensory neurons. Like odorants, many sapid chemicals, including most bitter-tasting substances, are highly lipophilic, their bitterness increasing with hydrophobicity¹³. Also, lipophilic sweeteners, such as saccharin, have much lower thresholds of detection than do hydrophilic carbohydrate sweeteners¹⁴. So the VEG protein could play a part in binding and presenting lipophilic gustatory molecules to the taste receptors. Alternatively, the VEG protein could be essential for effectively clearing the furrows of vallate and foliate papillae of lipophilic tastants, which otherwise would cause long-lasting taste sensations.

Whereas the binding of odorants, such as pyrazine, by OBP isolated from cow and rat has been demonstrated^{15,16}, a similar binding capacity for sapid chemicals by the VEG protein remains to be shown. The binding of odorants by the OBP and BG proteins seems to be nonspecific⁵. Likewise, the VEG protein is expected to recognize lipophilic substances nonspecifically, which would agree with the fact that many bitter compounds do not exhibit any structural relationship.

The visual, olfactory and, as reported here, gustatory sensory systems all depend on steps that are mediated by small lipophilic-molecule carrier proteins. The function of retinol-binding protein in visual perception, in which the transport of retinol to the photoreceptors involves the binding of retinol-RBP complexes to membrane-bound receptors, is the best-characterized example¹⁷. The presence of similar receptors for the olfactory binding proteins and the VEG protein of the gustatory system, and the physiological role of these proteins in the

binding of odorant and taste compounds, remains to be investigated. □

Received 2 October; accepted 23 November 1989.

1. Kinnamon, S. C. *Trends Neurosci.* **11**, 491-496 (1988).
2. Lancet, D. & Pace, U. *Trends biochem. Sci.* **12**, 63-66 (1987).
3. Pevsner, J., Reed, R. R., Feinstein, P. G. & Snyder, S. H. *Science* **241**, 336-339 (1988).
4. Cavaggoni, A., Sorbi, R. T., Keen, J. N., Pappin, D. J. C. & Findlay, J. B. C. *FEBS Lett.* **212**, 225-228 (1987).
5. Lee, K.-H., Wells, R. G. & Reed, R. R. *Science* **235**, 1053-1056 (1987).
6. Schofield, P. R. *Trends Neurosci.* **11**, 471-472 (1988).
7. Kinnamon, J. C. in *Neurobiology of Taste and Smell* (eds Finger, T. E. & Silver, W. L.) 277-297 (Wiley, New York, 1987).
8. Baumgartner, E. A. *Am. J. Anat.* **22**, 365-376 (1917).
9. Hand, A. R. *J. Cell Biol.* **44**, 340-353 (1970).
10. Godovac-Zimmermann, J. *Trends biochem. Sci.* **13**, 64-66 (1988).
11. Papiz, M. Z. *et al. Nature* **324**, 383-385 (1986).
12. Pevsner, J., Hwang, P. M., Sklar, P. B., Venable, J. C. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2383-2387 (1988).
13. Gardner, R. J. *Chem. Senses Flav.* **4**, 275-286 (1979).
14. Lawless, H. T. in *Neurobiology of Taste and Smell* (eds Finger, T. E. & Silver, W. L.) 401-420 (Wiley, New York, 1987).
15. Bignetti, E. *et al. Eur. J. Biochem.* **149**, 227-231 (1985).
16. Pevsner, J., Trifiletti, R. R., Strittmatter, S. M. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3050-3054 (1985).
17. Sivaprasadarao, A. & Findlay, J. B. C. *Biochem. J.* **255**, 561-569 (1988).
18. Schmale, H., Borowiak, B., Holtgreve-Grez, H. & Richter, D. *Eur. J. Biochem.* **182**, 621-627 (1989).
19. Lipman, D. J. & Pearson, W. R. *Science* **227**, 1435-1441 (1985).
20. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
21. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5463-5467 (1977).
22. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 6-13 (1983).
23. Higgins, G. A., Schmale, H., Bloom, F. E., Wilson, M. C. & Milner, R. J. *Proc. natn. Acad. Sci. U.S.A.* **86**, 2074-2078 (1989).

ACKNOWLEDGEMENTS. We thank Drs S. Fehr, S. Morley and D. Richter for reading the manuscript and Gitta Feyerabend for photographic expertise. This work was supported by Deutsche Forschungsgemeinschaft.

Activation of NMDA receptors induces dephosphorylation of DARPP-32 in rat striatal slices

Shelley Halpain, Jean-Antoine Girault* & Paul Greengard

Laboratory of Molecular and Cellular Neuroscience,
The Rockefeller University, New York, New York 10021, USA

IN the caudate-putamen the glutamatergic cortical input and the dopaminergic nigrostriatal input have opposite effects on the firing rate of striatal neurons¹⁻⁴. Although little is known of the biochemical mechanisms underlying this antagonism, one action of dopamine is to stimulate the cyclic AMP-dependent phosphorylation of DARPP-32 (dopamine and cAMP-regulated phosphoprotein, of relative molecular mass 32,000 (32K))⁵. This phosphorylation converts DARPP-32 from an inactive molecule into a potent inhibitor of protein phosphatase-1 (ref. 6). Here we show that activation of the NMDA (*N*-methyl-D-aspartate) subclass of glutamate receptors reverses the cAMP-stimulated phosphorylation of DARPP-32 in striatal slices through NMDA-induced dephosphorylation of DARPP-32. Thus, the antagonistic effects of dopamine and glutamate on the excitability of striatal neurons are reflected in antagonistic effects of these neurotransmitters on the state of phosphorylation of DARPP-32. Our results indicate that stimulation of NMDA receptors leads to the activation of a neuronal protein phosphatase, presumably the calcium-dependent phosphatase calcineurin, and show, in an intact cell preparation, that signal transduction in the nervous system can be mediated by protein dephosphorylation.

DARPP-32 is a neuronal phosphoprotein that is enriched in medium-sized spiny neurons of the caudate-putamen⁷. These cells receive dopaminergic input from the substantia nigra^{8,9} and glutamatergic input from the neocortex^{10,11}. The proposed

roles of dopamine and glutamate in regulating the phosphorylation and activity of DARPP-32 are illustrated in Fig. 1. Previous studies established that drugs that stimulate cAMP-dependent protein kinase activity in striatal cells, including dopamine and 8-bromo-cAMP, increase the phosphorylation of DARPP-32 (ref. 5). This phosphorylation occurs on a threonine residue that is phosphorylated *in vitro* by purified cAMP-dependent protein kinase¹². In addition, DARPP-32 is phosphorylated on serine residues by casein kinase II both *in vitro* and in striatal slices¹³. The threonine, but not the serine, phosphorylation site is dephosphorylated *in vitro* by calcineurin^{6,13}, a Ca²⁺/calmodulin-dependent protein phosphatase (protein phosphatase-2B)¹⁴. This observation indicated that extracellular signals that result in a rise in intracellular Ca²⁺ might, through activation of calcineurin, cause the dephosphorylation of DARPP-32 *in vivo*. One such extracellular signal might be the neurotransmitter glutamate, the best characterized receptor for which is the NMDA receptor. Activation of NMDA receptors leads to a Ca²⁺ influx both through the NMDA receptor channel itself^{15,16} and through voltage-sensitive Ca²⁺ channels activated by the resulting depolarization^{17,18}. Thus, activation of NMDA receptors could result in the activation of calcineurin and the dephosphorylation of DARPP-32 (ref. 19). Here, we use striatal slices from adult rat brain to investigate the effects of NMDA on the phosphorylation state of threonine and serine residues of DARPP-32.

Incubation of slices with [³²P]orthophosphate to label the cellular ATP pools resulted in a time-dependent incorporation of ³²P into DARPP-32. Incorporation was maximal after 60-90 min incubation, as measured by immunoprecipitation of DARPP-32 from the slices. Phosphoamino-acid analysis revealed that, under basal conditions, about 90% of the total ³²P in DARPP-32 was found on serine residues, with the remaining 10% on threonine residues (Fig. 2). Treatment of striatal slices for 5 min with forskolin (50 µM), an activator of adenylyl cyclase²⁰, resulted in a 2-4-fold increase in DARPP-32 threonine phosphorylation (Figs 2 and 3). Serine phosphorylation was not significantly affected by forskolin treatment (Figs 2 and 3).

The forskolin-stimulated phosphorylation of DARPP-32 was

* Present address: INSERM U114, Collège de France, 11 Place Marcelin Berthelot 75005, Paris, France.

abolished when NMDA (100 μ M) was added simultaneously with forskolin (Figs 2 and 3), indicating that NMDA antagonized the cAMP-dependent phosphorylation of DARPP-32. NMDA decreased threonine phosphorylation of DARPP-32, but had no significant effect on the level of serine phosphorylation (Figs 2 and 3). The NMDA effect seemed to require activation of the NMDA receptor channel, as both the competitive NMDA receptor antagonist AP5 (D-amino-5-phosphonopentanoic acid) and the noncompetitive NMDA receptor antagonist MK-801 completely blocked the effect of NMDA (Figs 2 and 3). Treatment with NMDA alone slightly decreased, and with AP5 alone slightly increased, the low basal level of threonine phosphorylation; treatment with forskolin plus AP5 gave results similar to treatment with forskolin plus NMDA plus AP5 (data not shown).

There are several ways in which NMDA receptor activation might result in an apparent decrease in DARPP-32 threonine phosphorylation. It could, for example, result in: selective proteolysis of DARPP-32 carrying phosphorylated threonine residues; inhibition of cAMP-dependent phosphorylation (by decreasing cAMP levels or by inhibiting cAMP-dependent protein kinase activity); or activation of a protein phosphatase that selectively dephosphorylates DARPP-32 threonine residues.

Treatment of slices with forskolin in the presence or absence of NMDA resulted in no change in the absolute levels of DARPP-32, according to an immunoblot analysis (Table 1). In considering the possible effects of NMDA on the cAMP-dependent phosphorylation of DARPP-32, two other results are important. First, Synapsin I, another substrate protein present in

TABLE 1 Total DARPP-32 concentration and cAMP-dependent protein phosphorylation

	Control	Forskolin	Forskolin + NMDA
Total DARPP-32 concentration† (c.p.m. μ g ⁻¹ homogenate)	118 $\pm$ 12	114 $\pm$ 7.3	124 $\pm$ 12
Synapsin I phosphorylation (site 1)‡ (arbitrary units)	2.99 $\pm$ 0.30	7.76 $\pm$ 1.20*	8.90 $\pm$ 0.66*
cAMP-dependent protein kinase activity§ (pmol min ⁻¹ mg ⁻¹)	2.76 $\pm$ 0.64	9.63 $\pm$ 2.64*	13.2 $\pm$ 3.42*

The concentration of forskolin was 50 μ M and that of NMDA was 100 μ M. Data correspond to the mean $\pm$ s.e.m. of 3–8 slices.

* Significantly different from control ($P < 0.05$, one-tail) by Kruskal-Wallis test and Mann-Whitney U posthoc analysis.

† Total DARPP-32 was assayed by immunoblot analysis²⁵ in striatal slices prepared as described in the legend to Fig. 2, but without the addition of [³²P]orthophosphate.

‡ Phosphorylation of the cAMP-dependent phosphorylation site (site 1) of synapsin I was determined by immunoprecipitation of synapsin I from prelabelled slices, as described for DARPP-32 in the legend to Fig. 2, using affinity purified antiserum against synapsin I (a gift of Dr A. J. Czernik). One-dimensional V-8 protease peptide mapping was performed as described²² to assay the effect of drug treatments specifically on site 1 of synapsin I.

§ Slices were prepared as described in the legend to Fig. 2, but without addition of ³²P-orthophosphate. Frozen slices were homogenized by sonication in 500 μ l ice cold buffer containing 5 mM HEPES, 2 mM EDTA, 1 mM EGTA, 2 mM β -mercaptoethanol, 1 mM phenylmethylsulphonyl fluoride, pH 7.4, at 4 °C. The homogenate was centrifuged at 250,000g for 10 min. An aliquot of the resulting supernatant was assayed for the amount of activated (cAMP-independent) form of cAMP-dependent protein kinase activity as described²⁶, using exogenous histone f2b as substrate.

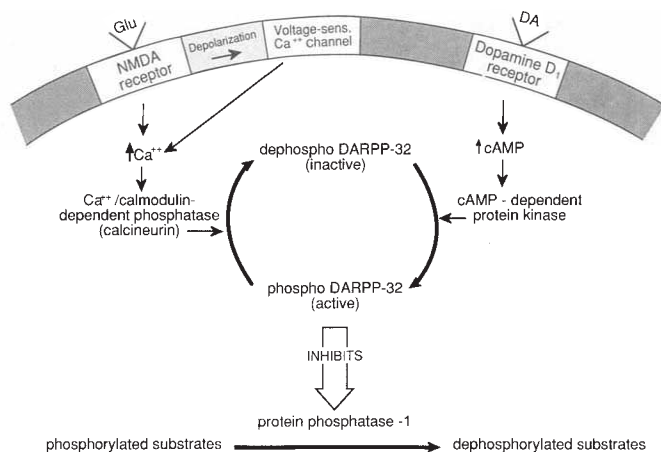


FIG. 1 Proposed mechanism for the dual regulation of the phosphorylation state and activity of DARPP-32 by two separate neurotransmitter systems which converge on DARPP-32-containing cells of the striatum (the medium-sized spiny neurons). Dopamine (DA) released from nerve terminals of nigrostriatal neurones activates the D1 subclass of dopamine receptors. These receptors stimulate adenylyl cyclase and thus increase levels of intracellular cAMP²⁷. The cAMP-dependent protein kinase becomes activated and phosphorylates DARPP-32 on a threonine residue¹², converting DARPP-32 into a potent inhibitor of protein phosphatase-1 (ref. 6). This phosphorylation and activation of DARPP-32 is reversed *in vitro* by the calcium/calmodulin-dependent phosphatase calcineurin⁶. Glutamate (Glu), the excitatory neurotransmitter released by corticostriatal nerve terminals^{10,11}, activates the NMDA subclass of glutamate receptors leading to an increase in the concentration of intracellular calcium, both by calcium entry through the NMDA receptor channel itself^{15,16}, and by calcium entry through voltage-sensitive calcium channels responding to NMDA-induced depolarization^{17,18}. In our model, this calcium entry leads to the activation of calcineurin. Calcineurin causes the dephosphorylation and inactivation of DARPP-32 (ref. 6), and the removal of inhibition of protein phosphatase-1. Protein phosphatase-1 is now free to dephosphorylate a variety of substrate proteins, such as the metabolic enzymes glycogen synthase and ribosomal protein S6, the synaptic vesicle proteins synapsins I and II, ion channels, and the structural proteins spectrin and microtubule-associated protein-2 (ref. 24).

striatal neurons²¹, which carries a phosphorylation site for cAMP-dependent protein kinase²², showed a twofold increase in phosphorylation in response to forskolin and this effect was not antagonized by co-incubating slices with NMDA (Table 1). By contrast, DARPP-32 immunoprecipitated from these same slices showed a clear antagonism of forskolin-induced phosphorylation by NMDA. Second, forskolin induced a three- to fourfold increase in the amount of the activated form of cAMP-dependent protein kinase in striatal slices, an increase which was not blocked by co-incubation with NMDA (Table 1). Together, these results indicate that activation of cAMP-dependent protein kinase by forskolin is not inhibited by treating slices with NMDA. That inhibition of adenylyl cyclase is not responsible for the NMDA effect is further supported by the observation that stimulation of DARPP-32 phosphorylation by 8-bromo-cAMP, which activates cAMP-dependent protein kinase directly, was also completely reversed by co-treatment with NMDA.

As neither proteolysis nor inhibition of phosphorylation can explain the effect of NMDA, it is likely that NMDA activates a protein phosphatase that selectively dephosphorylates DARPP-32 on threonine residues. Both protein phosphatase-2A and the Ca²⁺/calmodulin-dependent phosphatase calcineurin dephosphorylate DARPP-32 threonine residues *in vitro*, the latter being particularly effective⁶. On the other hand, phosphatase-2A, but not calcineurin, dephosphorylates DARPP-32 serine residues *in vitro*¹³. The lack of effect of NMDA on DARPP-32 serine phosphorylation strongly indicates that it is calcineurin which is responsible for the NMDA-induced dephosphorylation of DARPP-32 threonine residues. Moreover, NMDA receptor activation is known to raise intracellular calcium levels^{15,17}, and calcium in turn regulates calcineurin¹⁴. Although the calcium-dependence of the effect of NMDA could not be demonstrated because of generalized changes in protein phosphorylation in the complete absence of calcium (data not shown), we suggest that NMDA regulates the dephosphorylation of DARPP-32 through a mechanism involving calcineurin. Further studies are needed to determine whether other extracellular signals which also lead to increases in intracellular calcium have the same efficacy as NMDA in inducing DARPP-32 dephosphorylation. Although it seems probable that the

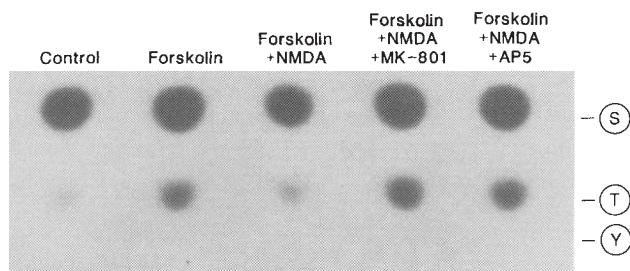


FIG. 2 NMDA reverses the stimulatory effect of forskolin on DARPP-32 threonine phosphorylation. This autoradiograph shows the phosphoamino-acid pattern seen in a typical experiment. Slices were incubated for 5 min in the presence of forskolin (50 μ M) in the absence or presence of NMDA (100 μ M) added simultaneously. The antagonists AP5 (100 μ M) and MK-801 ((+)-5-methyl-10,11-dihydro-5H-dibenzo[a,d]cyclohepten-5,10-imine; 10 μ M) were added 10 min before forskolin and NMDA. Control slices received no additions. The position of nonradioactive phosphoamino-acid standards is indicated on the right: S, phosphoserine; T, phosphothreonine, Y, phosphotyrosine.

METHODS. Striatal slices were prepared from adult female Sprague-Dawley rats as described²⁸, with the following modifications: A McIlwain tissue chopper was used to prepare 350 μ m-thick slices from freshly dissected striata. Dissection, cutting and preincubation (for 40 min) were carried out in a modified Krebs-bicarbonate buffer (phosphate-free) of the following composition: 124 mM NaCl, 4 mM KCl, 25 mM NaHCO₃, 1.5 mM MgSO₄, 15 μ M CaCl₂, 10 mM D-glucose, which was saturated with 95% O₂/5% CO₂ and adjusted to pH 7.3. At the end of this preincubation period, the buffer was removed and replaced with 0.5 ml of 'prelabelling buffer' (phosphate free, with no added magnesium) of the following composition: 124 mM NaCl, 4 mM KCl, 25 mM NaHCO₃, 1.5 mM CaCl₂, 10 mM D-glucose; 1 mCi [³²P]ortho-phosphate (specific activity, 100 mCi ml⁻¹, New England Nuclear). Slices were prelabelled for 90 min before application of various drugs. For each drug condition, incubations were carried out in triplicate. All compounds were diluted in prelabelling buffer. Forskolin was dissolved in ethanol, then diluted in prelabelling buffer, giving a final concentration of 0.5% ethanol in the slice medium. This concentration of ethanol was found to have no effect on DARPP-32 phosphorylation. Preliminary experiments showed that the effects of forskolin and NMDA were maximal within 5 min of drug application. Incubations were stopped by rapidly removing the medium with a pasteur pipette and immersing the tubes in liquid nitrogen. Slices were homogenized and protein concentration measured²⁸ and equalized for each sample. DARPP-32 was immunoprecipitated from slice homogenates (150 μ g protein per sample) as described previously²⁵, except that only a single cycle of immunoprecipitation was performed, using a mixture of affinity-purified monoclonal antibodies to purified DARPP-32 (a gift of Dr H. C. Hemmings Jr). The conditions used were previously shown to result in quantitative immunoprecipitation of DARPP-32 (ref. 25). Proteins were separated by electrophoresis on 12% SDS-polyacrylamide gels as described²², and gel pieces containing DARPP-32 were identified by alignment with the autoradiograph and excised. After measuring ³²P incorporation by counting Cerenkov radiation, gel pieces from triplicate samples were pooled and processed for phosphoamino-acid analysis²⁹. Following separation by thin-layer electrophoresis, the radioactive phosphoamino acids were visualized by autoradiography.

calcium signal caused by NMDA receptor activation would itself be sufficient to activate calcineurin and dephosphorylate DARPP-32, we cannot rule out the possibility that another substance released in response to NMDA contributes to this effect.

Calcineurin, like DARPP-32, is enriched in medium-sized spiny neurons of the striatum²³. Our results indicate that the glutamatergic input to these cells may function, in part, through receptor-dependent regulation of phosphatase activity. One target for this regulation is DARPP-32. Phosphorylation of DARPP-32 converts it into a potent inhibitor of protein phosphatase-1 (ref. 6) which, unlike calcineurin, has a broad substrate specificity²⁴. The significance of the regulation of protein phosphatase-1 in striatal neurons remains to be determined. However, *in vitro* studies in brain and other tissues have identified a variety of possible substrates for this enzyme, including synaptic vesicle proteins, ion channels, and structural proteins

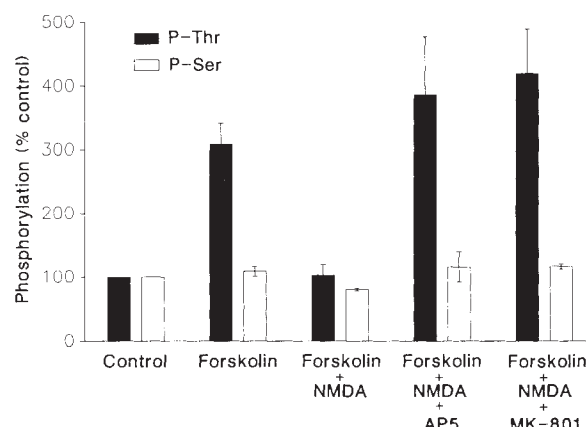


FIG. 3 Effects of forskolin, NMDA, AP5 and MK-801 on DARPP-32 phosphorylation. Results were averaged from 4–8 separate experiments. Data represent mean and standard error for each group. Abbreviations for the five treatment groups are as indicated in the legend to Fig. 2. The data are expressed as per cent of control. Results were analysed by Kruskal-Wallis test, followed by Mann-Whitney U posthoc analysis to probe for differences between groups. For the phosphothreonine site, forskolin treatment of slices, or forskolin plus NMDA in the presence of either NMDA antagonist, caused a significant increase in the amount of phosphorylation ($P < 0.05$). For phosphoserine, there were no statistically significant differences among the groups.

METHODS. Phosphoamino-acid analysis was performed as described in the legend to Fig. 2. The amount of ³²P in each phosphoamino acid was quantitated by liquid scintillation spectrometry. As the total amount of ³²P incorporation into proteins was found to vary between individual slices, a normalization procedure was used to correct for these differences. The amount of phosphorylated DARPP-32 in each slice was normalized by dividing the radioactivity incorporated into the immunoprecipitated DARPP-32 by the total ³²P incorporated into proteins as determined by trichloroacetic acid precipitation. The amounts of phosphorylated serine and threonine residues were calculated by multiplying this value by the ratio of [³²P]phosphoserine or [³²P]phosphothreonine to ([³²P]phosphoserine + [³²P]phosphothreonine), respectively, determined by phosphoamino-acid analysis. Other methods are described in the legend to Fig. 2.

(see ref. 24 for review). Dephosphorylation of DARPP-32 should result in increased activity of protein phosphatase-1 and therefore in the dephosphorylation of these protein phosphatase-1 substrates. Figure 1 provides a schematic representation of how the phosphorylation state and functional activity of DARPP-32 can be regulated by two distinct but converging neurotransmitter systems (in this case, glutamate and dopamine) involving two distinct second messengers (calcium and cAMP). Our data provide a specific example of how signals that act through calcium can attenuate signals that act through cAMP^{14,24}. □

Received 4 August; accepted 14 November 1989.

- Johnson, S. W., Palmer, M. R. & Freedman, R. *Neuropharmacology* **22**, 843–851 (1983).
- Hirata, K., Yim, C. Y. & Mogenson, G. J. *Brain Res.* **321**, 1–8 (1984).
- Calabresi, P., Mercuri, N., Stanzione, P., Stefani, A. & Bernardi, G. *Neuroscience* **20**, 757–771 (1987).
- Siggins, G. R. in *Psychopharmacology: A Generation of Progress* (eds Lipton, M. A., DiMascio, A. & Killam, K. F.) 143–157 (Raven, New York, 1978).
- Walaas, S. I., Aswad, D. W. & Greengard, P. *Nature* **301**, 69–71 (1983).
- Hemmings, H. C. Jr, Greengard, P., Tung, H. Y. L. & Cohen, P. *Nature* **310**, 503–505 (1984).
- Quimet, C. C., Miller, P. E., Hemmings, H. C. Jr, Walaas, S. I. & Greengard, P. *J. Neurosci.* **4**, 111–124 (1984).
- Pickel, V. M., Beckley, S. C., Joh, T. H. & Reis, D. J. *Brain Res.* **225**, 373–385 (1981).
- Freund, T. F., Powell, J. F. & Smith, A. D. *Neuroscience* **13**, 1189–1215 (1984).
- Fonnum, F., Storm-Mathisen, J. & Divac, I. *Neuroscience* **6**, 863–873 (1981).
- Girault, J. A. et al. *J. Neurochem.* **47**, 98–106 (1986).
- Hemmings, H. C. Jr, Nairn, A. C., Aswad, D. W. & Greengard, P. *J. Neurosci.* **4**, 99–110 (1984).
- Girault, J. A., Hemmings, H. C. Jr, Williams, K. R., Nairn, A. C. & Greengard, P. *J. Biol. Chem.* (in the press).
- Klee, C. B. & Cohen, P. in *Molecular Aspects of Cellular Regulation* (eds Cohen, P. & Klee, C. B.) 225–248 (Elsevier, Amsterdam, 1988).
- MacDermott, A. B., Mayer, M. L., Westbrook, G. L., Smith, S. J. & Barker, J. L. *Nature* **321**, 519–522 (1986).
- Jahr, C. E. & Stevens, C. F. *Nature* **325**, 522–525 (1987).
- MacDermott, A. B. & Dale, N. *Trends Neurosci.* **10**, 280–284 (1987).
- Hockberger, P. E., Tseng, H.-Y. & Connor, J. A. *J. Neurosci.* **9**, 2272–2284 (1989).
- Hemmings, H. C. Jr, Walaas, S. I., Quimet, C. C. & Greengard, P. in *Receptor Biochemistry and Methodology, Vol. 9, Structure and Function of Dopamine Receptors* (eds Creese, I. & Fraser, C. M.) 115–151 (New York, 1987).

20. Seamon, K. B. & Daly, J. W. *Adv. Cyclic Nucleotide Protein Phosphorylation Res.* **20**, 1-149 (1986).
21. Walaas, S. I., Nairn, A. C. & Greengard, P. *J. Neurosci.* **3**, 291-301 (1983).
22. Huttner, W. B. & Greengard, P. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5402-5406 (1979).
23. Goto, S., Matsukado, Y., Miyamoto, E. & Yamada, M. *Neuroscience* **22**, 189-201 (1987).
24. Cohen, P. *Ann. Rev. Biochem.* **58**, 453-508 (1989).
25. Hemmings, H. C. Jr. & Greengard, P. *J. Neurosci.* **6**, 1469-1481 (1986).
26. Witt, J. J. & Roskoski, R. Jr. *Analyt. Biochem.* **66**, 253-258 (1975).
27. Kebedian, J. W. & Calne, D. B. *Nature* **277**, 93-96 (1979).
28. Hemmings, H. C. Jr., Girault, J. A., Williams, K. R., LoPresti, M. B. & Greengard, P. *J. biol. Chem.* **264**, 7726-7733 (1989).
29. Nairn, A. C. & Greengard, P. *J. biol. Chem.* **262**, 7273-7281 (1987).

ACKNOWLEDGEMENTS. This work was supported by a grant from the NIMH. J.-A.G. was on leave-of-absence from INSERM U114.

Coexpression of two distinct muscle acetylcholine receptor α -subunits during development

Deborah S. Hartman & Toni Claudio^{*†}

Department of Biology, Yale University, 219 Prospect St., New Haven, Connecticut 06511, USA

^{*}Department of Cellular and Molecular Physiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

[†]To whom correspondence should be addressed

THE nicotinic acetylcholine receptor is a ligand-gated channel that mediates signalling at the vertebrate neuromuscular junction. It is a pentameric complex of four different subunits, assembled with a stoichiometry of $\alpha_2\beta\gamma\delta$. Muscle-like α -subunits have been cloned from *Torpedo*, mouse, calf, rat, chicken, human and *Xenopus*, and only a single α -subunit complementary DNA from each species has been detected (reviewed in ref. 1). We report here the cloning and characterization of a second muscle α -subunit cDNA from *Xenopus*, and show that this and a previously reported *Xenopus* α -subunit cDNA² are encoded by distinct genes. The novel α -subunit reported here is expressed uniquely in oocytes;

but both types of α -subunit are coexpressed throughout muscle development. This latter observation indicates that the expression of these two α -subunits is different from a previously reported developmental 'subunit-switch' mechanism used to generate channel diversity.

The four different subunits of the muscle-like nicotinic acetylcholine receptors (AChRs) are quite similar within a species, displaying pairwise identities of ~35-50%. Of the AChR subunits, α -subunits are the most highly conserved among species, and they can be unambiguously identified by the presence of adjacent Cys residues at positions 192 and 193, which are thought to be located at or near the ACh-binding site³. Both muscle-like α -subunits (α_1) and neuronal α -subunits (designated by ' α ' followed by a number >1) possess such adjacent Cys residues; only mature muscle-like α -subunits, however, are composed of 437 amino acids¹. Until now, data have indicated that there is only a single α -subunit gene within a species. In apparent contrast to this result, pharmacological evidence demonstrates nonequivalence in the binding of the two α -subunits by both agonists and antagonists⁴. These findings are usually reconciled by noting that the micro-environment of each α -subunit is different, because each is flanked by a unique pair of non- α -subunits⁵. If two different α -subunits are coexpressed and able to assemble into the same pentamer, then differences in the binding properties of the α -subunits might be expected, as well as differences in other channel properties.

We screened a *Xenopus* embryonic cDNA library⁶ with a

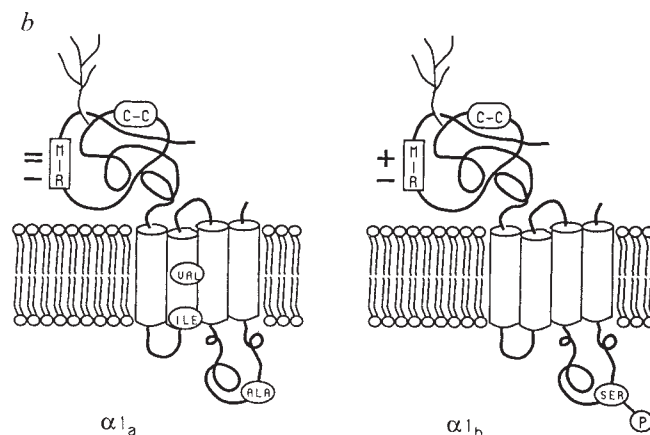
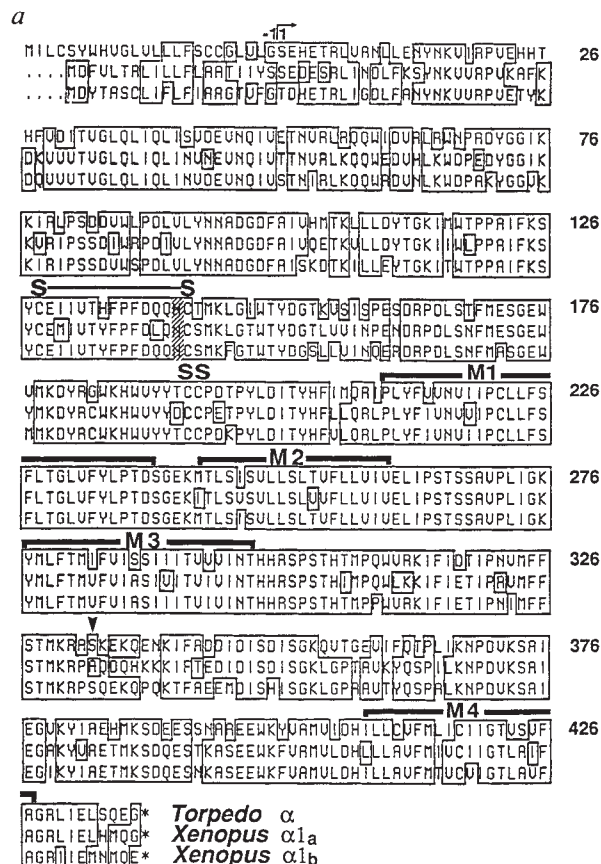


FIG. 1 **a**, Deduced amino-acid sequences of *Torpedo* α -subunit, *Xenopus* α_1 _a and *Xenopus* α_1 _b. Boxes indicate amino-acid identities among the three sequences. S, predicted disulphide linkages; the potential N-linked glycosylation site at position 141 is indicated by hatched boxes, and the potential phosphorylation site at serine 333 is marked with an arrowhead. Four putative transmembrane domains (M1-M4) are identified. Clone α_1 _a has an open reading frame coding for 457 amino acids with a predicted 20 amino-acid signal sequence. The deduced amino-acid sequence of α_1 _a is 89% similar to that of α_1 _b. Subunit α_1 _a is 85%, 63%, 57% and 56% similar to *Torpedo* α -, β -, γ - and δ -subunits, respectively. The nucleotide sequence of α_1 _a is available in the EMBL, GenBank and DDBJ Nucleotide Sequence databases under the accession number X17244. **b**, Subunit folding models comparing α_1 _a and α_1 _b structures. Two nonpolar residues (valine, isoleucine) in M2 of α_1 _a correspond to polar residues in all other muscle-like α -subunits. A putative phosphorylation site (Ser 333) is present in all muscle-like α -subunits except α_1 _a. The main immunogenic region (MIR) of α_1 _a is nearly identical to the MIR of *Torpedo* and mammalian α -subunits, but there is a difference of three unit charges between α_1 _a and α_1 _b in this region (residues 67-76)¹⁹. (C-C) indicates adjacent cysteines residues. Both subunits are shown glycosylated.

METHODS. A full-length *Torpedo* α -subunit cDNA probe⁷ identified four α_1 _a clones in 500,000 recombinant phage of a stage-17 *Xenopus* cDNA library⁶. Filters were hybridized at 48 °C in aqueous solution²⁰. In a second screen, four additional α_1 _a clones and three α_1 _b clones were identified from 500,000 phage using a partial α_1 _b cDNA probe². Hybridizations were performed in 50% formamide solution²⁰ and filters were washed in 0.5×SSC, 0.1% SDS at 45 °C. Sequencing reactions were done on single-stranded M13 templates using Klenow fragment (Boehringer) and dideoxy nucleotides (Pharmacia).

full-length *Torpedo* α -subunit probe⁷ and isolated four clones. These cDNAs encode a protein with a predicted signal peptide of 20 amino acids and a mature protein of 437 amino acids which contains two adjacent Cys residues located at positions 192 and 193 (Fig. 1). Because of these last two structural features, we tentatively identified these new clones as muscle-like α -subunits (designated $\alpha 1_a$). We compared the sequences of the

Xenopus $\alpha 1_a$ subunit, the previously reported *Xenopus* α -subunit² (designated $\alpha 1_b$, by agreement with the authors², because it is expressed later in development than $\alpha 1_a$ —see below), and the muscle-like α -subunits from *Torpedo*, bovine, human and mouse. The $\alpha 1_a$ and $\alpha 1_b$ subunits are as divergent from each other (89% similar) as either is from the mammalian or *Torpedo* (85–89%) α -subunit, and amino-acid differences are present throughout the two sequences; this indicates that $\alpha 1_a$ and $\alpha 1_b$ subunits are encoded in separate genes.

Genomic Southern blots confirmed the presence of two distinct α -subunit genes (Fig. 2a). Each gene is present in a single copy per genome and each is encoded by several exons (Fig. 2b). The longest stretch of identity between $\alpha 1_a$ and $\alpha 1_b$ is only 30 base pairs, and most DNA fragments from the two cDNAs do not cross-hybridize on Southern blots. If multiple α -subunit genes exist in organisms other than *Xenopus*, then they may have gone undetected if their DNAs have diverged as much as those of the two *Xenopus* α -subunit genes.

To determine whether the α -subunit genes encode functional polypeptides, we engineered $\alpha 1_a$ and $\alpha 1_b$ cDNAs into SP6 and mammalian expression vectors. Complementary RNA transcripts made *in vitro* and expressed in an *in vitro* translation system produce proteins of relative molecular mass $\sim 40,000$ (M_r 40 K) which are recognized by α -subunit-specific antisera⁸. Larger species are produced in the presence of canine pancreas microsomes (Fig. 3), consistent with the addition of one unit of *N*-linked oligosaccharide per polypeptide (the number observed or predicted for all muscle-like α -subunits¹). In addition, current

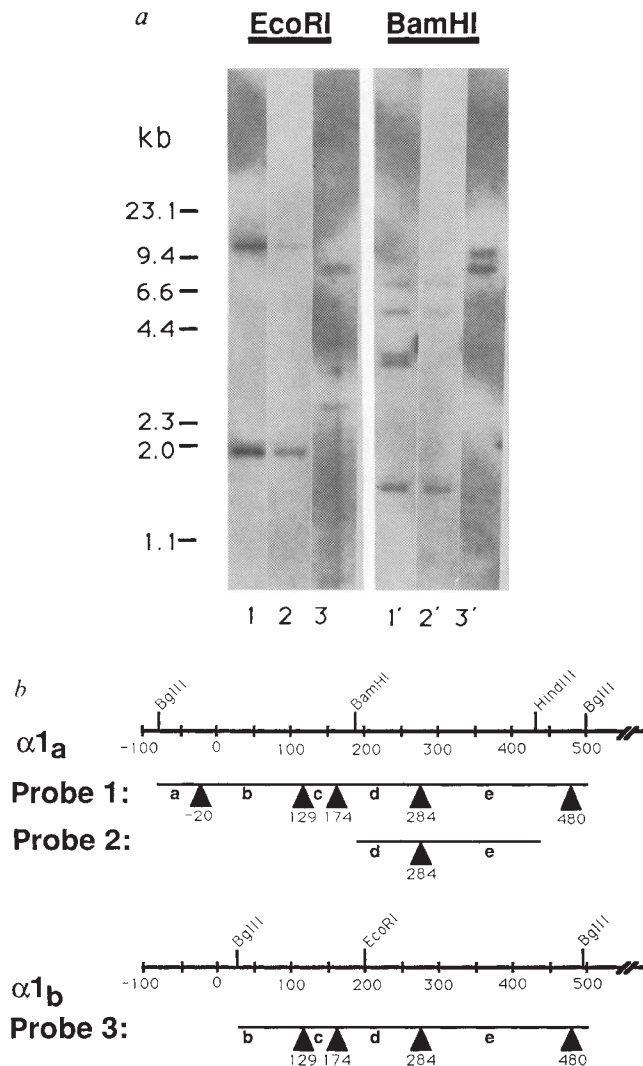


FIG. 2 Southern analysis of *Xenopus* genomic DNA. *a*, Genomic DNA isolated from liver tissue was digested with five restriction endonucleases and hybridized to two $\alpha 1_a$ probes (lanes 1, 1', 2, and 2') and an $\alpha 1_b$ probe (lanes 3 and 3'). In all digests, restriction fragments identified by $\alpha 1_a$ probes were distinct from those identified by $\alpha 1_b$ probes, indicating the presence of two different genes. Intensities of hybridized fragments compared with those of control lanes with known amounts of $\alpha 1_a$ and $\alpha 1_b$ DNA, indicate that each gene is present in a single copy per genome. Sizes of M_r markers are shown (kb, kilobases). *b*, Partial restriction maps of $\alpha 1_a$ and $\alpha 1_b$ genes are shown, with probe fragments indicated below. Multiple probes were used to locate intron positions. A minimum of two introns are identified in the $\alpha 1_a$ gene, one 5' and the other 3' to the internal *Bam*HI site. At least one intron is identified in the $\alpha 1_b$ gene by each of four different restriction digests (data not shown). Muscle AChR α -subunit genes have been isolated from human²¹ and chicken²², and the location of intron-exon borders is conserved between the two species. Predicted intron-exon borders are indicated by black triangles.

METHODS. Genomic DNA was isolated from adult *Xenopus* liver tissue using Proteinase K-SDS²⁰. DNA was digested with *Eco*RI, *Sac*I, *Xba*I, *Hind*III and *Bam*HI enzymes, electrophoresed on two identical 0.8% gels, and transferred to BioTrace RP membranes (Gelman) by alkaline transfer. Random-primed DNA probes (Amersham) were hybridized under stringent conditions (50% formamide, 42 °C) and washed in 0.1 × SSC, 0.1% SDS at 68 °C.

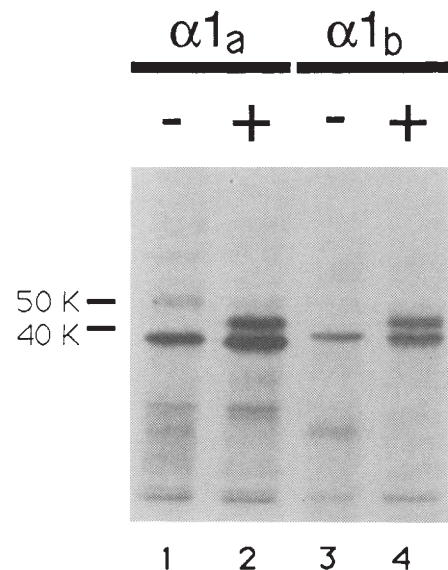


FIG. 3 *In vitro* translation and immunoprecipitation of $\alpha 1_a$ and $\alpha 1_b$ cDNA and encoded proteins, respectively. Complementary RNA $\alpha 1_a$ and $\alpha 1_b$ transcripts were made *in vitro* and translated in rabbit reticulocyte lysates in the presence of [³⁵S]methionine. Although both α -subunit cDNAs encode 457 amino acids, $\alpha 1_a$ (calculated M_r of 52.5 K) has a slightly faster electrophoretic mobility than $\alpha 1_b$ (52.1 K); (lanes 1 and 3, respectively). The total charge per molecule is -4 for $\alpha 1_a$ and -3 for $\alpha 1_b$, which could explain the observed electrophoretic mobility differences. Larger M_r species of both α -subunits are made in the presence of canine pancreas microsomes (lanes 2 and 4), consistent with the addition of one unit of polysaccharide per polypeptide. Both processed and unprocessed forms of the proteins are recognized by polyclonal antisera directed against the *Torpedo* α -subunit⁸. **METHODS.** Complementary $\alpha 1_a$ DNAs were subcloned into plasmid pGEM-4 (Promega), and RNA was transcribed *in vitro* using a SP6 vector system²³. Transcripts were translated in an RNA-free rabbit reticulocyte lysate (Promega) with or without canine pancreas microsomes (Amersham) according to manufacturer's instructions. Translation lysates were solubilized in a buffer containing 0.5% Nonidet P-40 and immunoprecipitated as previously described¹¹. Proteins were separated on 10% SDS polyacrylamide gels.

experiments indicate that $\alpha 1_a$ and $\alpha 1_b$ form high-affinity α -bungarotoxin-binding AChR complexes when coexpressed with either *Torpedo* or mouse β -, γ -, δ -subunits stably expressed in NIH3T3 cells (Y. Wang and T.C., unpublished results).

Previously, two γ -AChR-subunits have been reported. In bovine⁹, rat¹⁰ and *Xenopus*², γ -subunit messenger RNA transcripts are detected early but not late in development. Transcripts for γ -like (ϵ) subunits appear late in development in cow and rat, and are not detectable at earlier stages. To determine whether there is a similar developmental switch between *Xenopus* $\alpha 1_a$ and $\alpha 1_b$ subunit transcripts, we performed RNase protection assays on RNA isolated from oocytes, embryos, and embryonic and adult muscle tissues (Fig. 4a). Both $\alpha 1_a$ and $\alpha 1_b$ transcripts are clearly present from early myoblast differentiation to the formation of mature neuromuscular junctions in myotomal (tail)

muscle (stage 14–50) and both are present in isolated embryonic and adult muscle tissues (myotomal, interhyoid, leg). The expression of *Xenopus* $\alpha 1_a$ and $\alpha 1_b$ transcripts is therefore quite dissimilar from that of γ - and ϵ -subunits. Northern analysis showed that $\alpha 1_a$ is not expressed in brain or liver (Fig. 4d). It is interesting to note that $\alpha 1_a$ transcripts are expressed in oocytes (Fig. 4a). Expression of an endogenous AChR subunit might be problematic in this system when trying to express exogenous muscle-like or neuronal AChR subunits, as well as subunits from other ligand-gated channels with similar structures.

Although $\alpha 1_a$ and $\alpha 1_b$ transcript levels are not equivalent at all embryonic stages (Fig. 4c), both are expressed in the same muscle tissues and at mostly the same developmental stages. The functional significance of this coexpression has not yet been determined, but it indicates that multiple channel types could

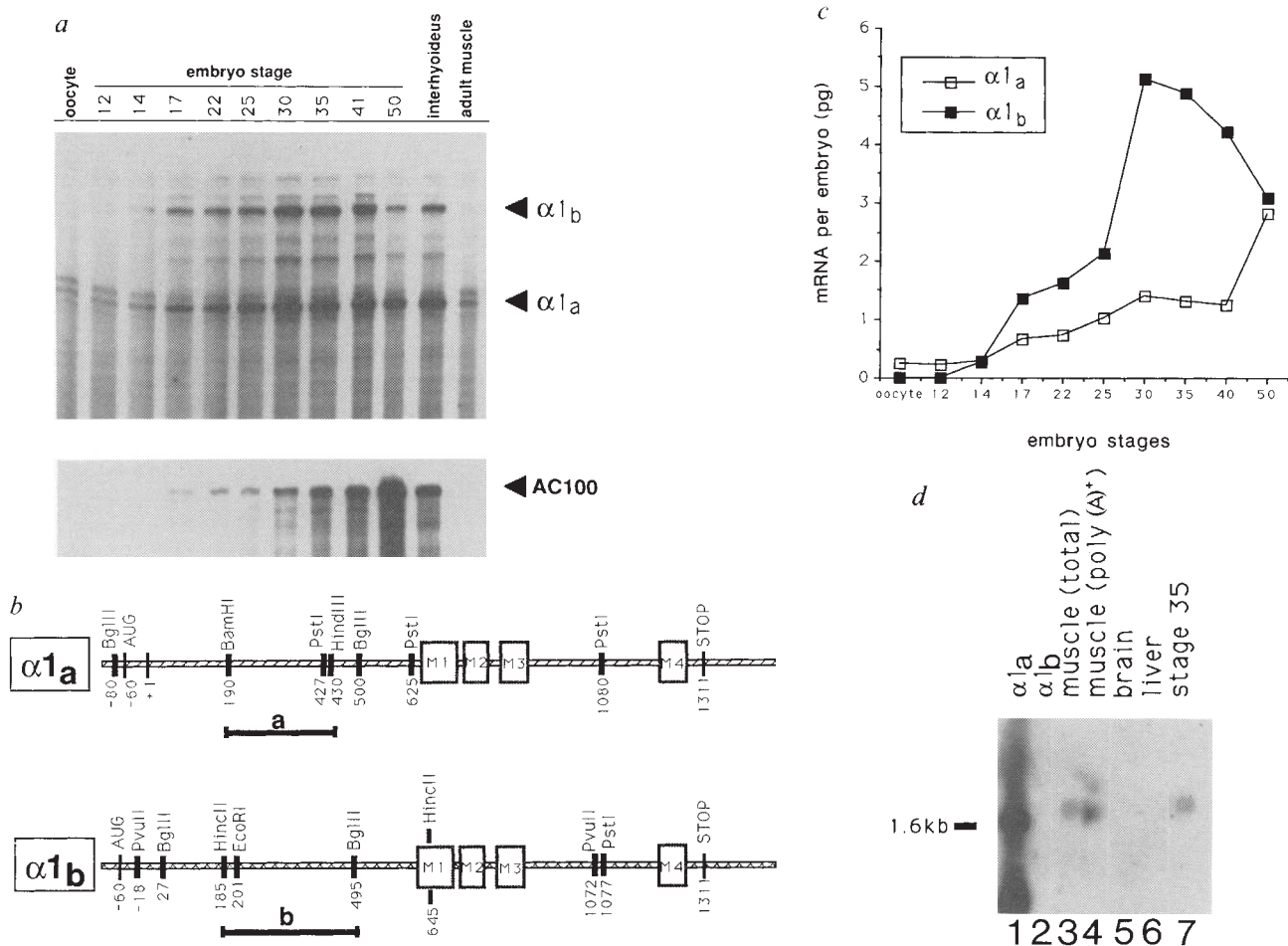


FIG. 4 Coexpression of $\alpha 1_a$ and $\alpha 1_b$ transcripts during *Xenopus* development and in mature muscle tissues. **a**, RNase protection analysis of $\alpha 1_a$, $\alpha 1_b$, and actin transcripts expressed during development. Sizes of the protected $\alpha 1_a$, $\alpha 1_b$, and actin probe fragments are indicated by arrowheads. **b**, Probes for $\alpha 1_a$ and $\alpha 1_b$ transcripts used in RNase-protection and northern blot analyses are identified below their respective cDNA restriction maps. **c**, Expression levels of $\alpha 1_a$ and $\alpha 1_b$ change during development. In the oocyte and in stage-12 embryos, only $\alpha 1_a$ transcripts are detected. Transcripts for $\alpha 1_b$ are 2–5-fold more abundant than $\alpha 1_a$ transcripts through myotomal muscle synaptogenesis in embryos (stages 17–40), but $\alpha 1_a$ and $\alpha 1_b$ transcripts are expressed at nearly equal levels in mature muscles (stage-50 myotomal, interhyoid, leg). **d**, Northern analysis of $\alpha 1_a$ expression in embryonic and adult tissue. A single $\alpha 1_a$ transcript is detected in adult leg muscle and in stage-35 embryo tails, but no hybridization is detected in liver or brain tissue. Lanes 1 and 2 contain 10 pg $\alpha 1_a$ and $\alpha 1_b$ sense RNA, respectively, made *in vitro* using SP6 polymerase (see legend to Fig. 3); lane 3, 10 μ g total RNA from adult leg muscle; lane 4, 1 μ g poly(A)⁺ RNA from adult leg muscle; lanes 5–7, 10 μ g total RNA from brain, liver and stage-35 tails, respectively.

METHODS. RNase protection: *Xenopus* embryos were obtained from Nasco,

and staged according to previously defined criteria²⁴. Tails were isolated from stage-30–50 embryos, and interhyoid muscle was enzymatically isolated from stage-50 embryos. Myotomal muscle from stage-50 embryo tails was dissected away from central nervous system and epidermal tissue. Samples were homogenized in 25 μ l GSCN (5 M), 100 mM EDTA per embryo, and hybridized to ³²P-labelled antisense RNA probes at 27 °C (ref. 25). The $\alpha 1_a$ probe was 280 bp (250 bp protected), and the $\alpha 1_b$ probe was 340 bp (315 bp protected). The SP64-*Xenopus* cardiac actin AC100 construct²⁶ was linearized with *PvuII*, giving a probe of 591 nucleotides (550 bp protected). Samples were analysed on 5% polyacrylamide–50% urea gels. Amount of mRNA per embryo was quantified by densitometric scanning of autoradiograph bands using a Kodak Visage 2000 after correcting for probe intensities by serial dilutions of sense RNA to calibrate levels of protected transcript. Northern blots: Poly(A)⁺ and total RNAs were isolated from adult tissues using standard procedures²⁰. RNA was electrophoresed on formaldehyde agarose gels and electroblotted to Zetabind filters (AMF Cuno). Filters were hybridized in 50% formamide, 0.25 M sodium phosphate, pH 7, 0.25 M NaCl, 7% SDS, and 1 mM EDTA at 68 °C, and washed in 0.5 $\times$ SSC, 0.1% SDS at 68 °C. Antisense RNA probes were made according to published procedures²³. Sense-strand RNAs made *in vitro* were used as *M*_i markers.

be expressed if the two α -subunits were present in the same pentamer. We have established rat L6 muscle cell lines that stably express *Torpedo* α -subunits¹¹, and have demonstrated that single AChR pentamers can be expressed in these cell lines which are composed of rat β -, γ - and δ -subunits together with one *Torpedo* α -subunit and one rat α -subunit¹². It is possible, therefore, that *Xenopus* AChR pentamers composed of $\alpha 1_a$ and $\alpha 1_b$ subunits also form. If so, four channel types would be predicted: pentamers composed of two $\alpha 1_a$ subunits, two $\alpha 1_b$ subunits, or one of each, located in two unique positions.

Electrophysiological recordings from *Xenopus* myotomal muscle demonstrate the presence of at least three AChR channel types

(25, 40 and 60 pS channels)¹³ with similar diversity found in mammalian muscle AChRs¹⁴. Evidence indicates that the transmembrane domain M2 of each subunit forms the lining of the channel¹⁵⁻¹⁸. Of interest is that the M2 domain of $\alpha 1_a$ has two fewer polar residues than the M2 domains of other α -subunits (including $\alpha 1_b$), which could result in different channel conductances. One other unique structural feature of $\alpha 1_a$ is the lack of the single potential cyclic AMP-dependent protein-kinase phosphorylation site located at position 333 in all other cloned α -subunits¹. The presence or absence of this site could provide a mechanism for distinguishing between or regulating the expression of $\alpha 1_a$ and $\alpha 1_b$ subunits during development. $\square$

Received 15 August; accepted 16 November 1989.

- Claudio, T. in *Frontiers of Molecular Biology, Molecular Neurobiology* (eds Glover, D. & Hames, D.) 9-88 (IRL, London, 1989).
- Baldwin, T. J., Yoshihara, C. M., Blackmer, K., Kintner, C. R. & Burden, S. J. *J. Cell Biol.* **106**, 469-478 (1988).
- Kao, P. N. et al. *J. Biol. Chem.* **259**, 11662-11665 (1984).
- Sine, S. M. & Taylor, P. J. *J. Biol. Chem.* **255**, 10144-10156 (1980).
- Karlin, A. *Nature* **329**, 286-287 (1987).
- Kintner, C. R. & Melton, D. A. *Development* **99**, 311-325 (1987).
- Claudio, T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 5967-5971 (1987).
- Claudio, T. & Raftery, M. A. *Archs Biochem. Biophys.* **181**, 484-489 (1977).
- Takai, T. et al. *Nature* **315**, 761-764 (1985).
- Witzemann, V., Barg, B., Criado, M., Stein, E. & Sakmann, B. *FEBS Lett.* **242**, 419-424 (1989).
- Claudio, T. et al. *J. Cell Biol.* **108**, 2277-2290 (1989).
- Paulson, H. L. & Claudio, T. *Soc. Neurosci. Abstr.* **14**, 1045 (1988).
- Owens, J. & Kullberg, R. J. *J. Neurosci.* **9**, 2575-2580 (1989).
- Brehm, P. & Henderson, L. *Dev. Biol.* **129**, 1-11 (1988).
- Hucho, F., Oberthur, W. & Lottspeich, F. *FEBS Lett.* **205**, 137-142 (1989).

- Giraudat, J., Dennis, M., Heidmann, T., Chang, J.-Y. & Changeux, J.-P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2719-2723 (1986).
- Imoto, K. et al. *Nature* **335**, 645-648 (1988).
- Leonard, R. J., Labarca, C. G., Charney, N. & Lester, H. A. *Science* **242**, 1578-1581 (1988).
- Tzartos, S. J., Kokla, A., Walgrave, S. L. & Conti-Tronconi, B. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 2899-2903 (1988).
- Maniatis, T., Fritsch, E. F., & Sambrook, J. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).
- Noda, M. et al. *Nature* **305**, 818-823 (1983).
- Nef, P., Oneyser, C., Barkas, T. & Ballivet, M. in *NATO ASI Series, Vol H3* (ed. Maelicke, A.) 417-422 (Springer, Berlin, 1986).
- Krieg, P. A. & Melton, D. A. in *Meth. Enzym.* **155**, 397-415 (1987).
- Nieuwkoop, P. D. & Faber, J. *Normal Table of Xenopus laevis* (Daudin, Amsterdam, 1967).
- Thompson, J. & Gillespie, D. *Analyt. Biochem.* **163**, 281-291 (1987).
- Dworkin-Rastl, E., Kelly, D. B. & Dworkin, M. B. *J. Embryol. exp. Morph.* **91**, 153-168 (1986).

ACKNOWLEDGEMENTS. We thank Steve Burden for the gift of $\alpha 1_a$ cDNAs; Jesse Owens for interhyoid muscle and myotomal muscle from stage 50 *Xenopus* embryos, and Tom Sargent for the *Xenopus* cardiac actin construct AC100.

Modulation of ATP-sensitive K^+ channels in skeletal muscle by intracellular protons

N. W. Davies

Department of Physiology, University of Leicester, PO Box 138, Leicester LE1 9HN, UK

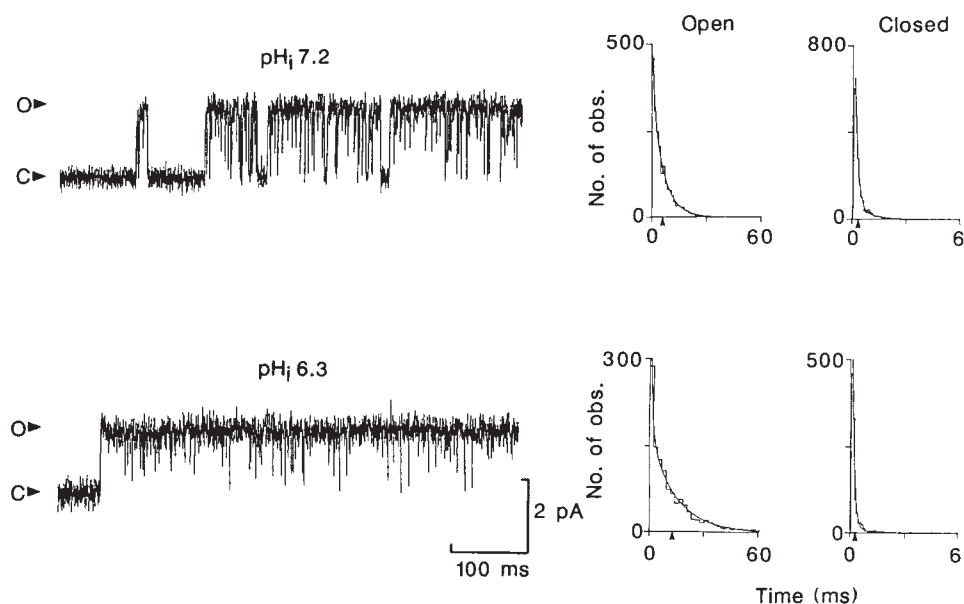
SINCE their discovery in cardiac muscle¹, ATP-sensitive K^+ (K_{ATP}) channels have been identified in pancreatic β -cells², skeletal muscle³, smooth muscle⁴ and central neurons⁵. The activity of K_{ATP} channels is inhibited by the presence of cytosolic ATP. Their wide distribution indicates that they could have important physiological roles that may vary between tissues. In muscle cells the role of K^+ channels is to control membrane excitability and the duration of the action potential. In anoxic cardiac ventricular muscle K_{ATP} channels are believed to be responsible for shortening the action potential⁶, and it has been proposed that a fall in ATP concentration during metabolic exhaustion increases the activity of K_{ATP} channels in skeletal muscle⁷, which may reduce excitability. But the intracellular concentration of ATP in muscle is buffered by creatine phosphate to 5-10 mM, and changes little, even during sustained activity⁸. This concentration is much higher than the intracellular ATP concentration required to half block the K_{ATP} -channel current in either cardiac muscle (0.1 mM)¹ or skeletal muscle (0.14 mM)⁹, indicating that the open-state probability of K_{ATP} channels is normally very low in intact muscle. So it is likely that some additional means of regulating the activity of K_{ATP} channels exists, such as the binding of nucleotides other than ATP¹⁰⁻¹². Here I present evidence that a decrease in intracellular pH (pH_i) markedly reduces the inhibitory effect of ATP on these channels in excised patches from frog skeletal muscle. Because sustained muscular activity can decrease pH_i by almost 1 unit^{13,14} in the range at which K_{ATP} channels are most sensitive to pH_i , it is likely that the activity of these channels in skeletal muscle is regulated by intracellular protons under physiological conditions.

Experiments were performed at room temperature (20-23 °C) on inside-out patches¹⁵ isolated from frog sarcolemmal vesicles prepared by KCl-loading and enzymatic treatment¹⁶. Patch electrodes were coated with Sylgard, fire-polished and filled with 10 mM K^+ -110 mM Na^+ solutions at pH 7.2. Test (intracellular) solutions, which were applied to the cytoplasmic side of the patch, all contained 120 mM K^+ , 5 mM EGTA to chelate Ca^{2+} , and 10 mM buffer (HEPES, PIPES or MES): $MgCl_2$ and K_2 -ATP or Na_2 -ATP were added as necessary and the pH readjusted. Under these conditions, K_{ATP} -channel activity was readily observed and contamination from the activities of other channel types was negligible. For kinetic analysis, only recordings containing one active channel were used.

In the absence of ATP, a decrease in pH_i has a significant effect on the K_{ATP} -channel kinetics. The main observation was an increase in the mean open time of the channel as pH_i was decreased. The closed time distribution revealed an increase in the number of short closings, and a decrease in the time constant of this component. The open-state probability P_{open} of these channels is determined mainly by the existence of long closed durations, and despite the marked change in open time distribution, P_{open} was only slightly altered by changing pH_i in the absence of ATP (see Fig. 3). The kinetic effect of changing pH_i on long closed durations was not examined. In five patches at a holding potential of -3 mV, the mean open times ($\pm$ s.e.m.), corrected for missed closings (see Fig. 1 legend), were 2.27 ± 0.40 , 4.53 ± 0.81 , 7.43 ± 0.21 and 7.50 ± 0.33 ms at pH_i 8.0, 7.2, 6.3 and 5.7, respectively. The effect of reducing pH_i from 7.2 to 6.3 on a single K_{ATP} channel in the absence of ATP at a holding potential of -3 mV is shown in Fig. 1. There was also a progressive reduction in single-channel amplitude on decreasing pH_i , the means ($\pm$ s.e.m.) of five patches being 1.86 ± 0.05 , 1.79 ± 0.04 , 1.49 ± 0.03 and 1.48 ± 0.03 pA at pH_i 8.0, 7.2, 6.3 and 5.7, respectively.

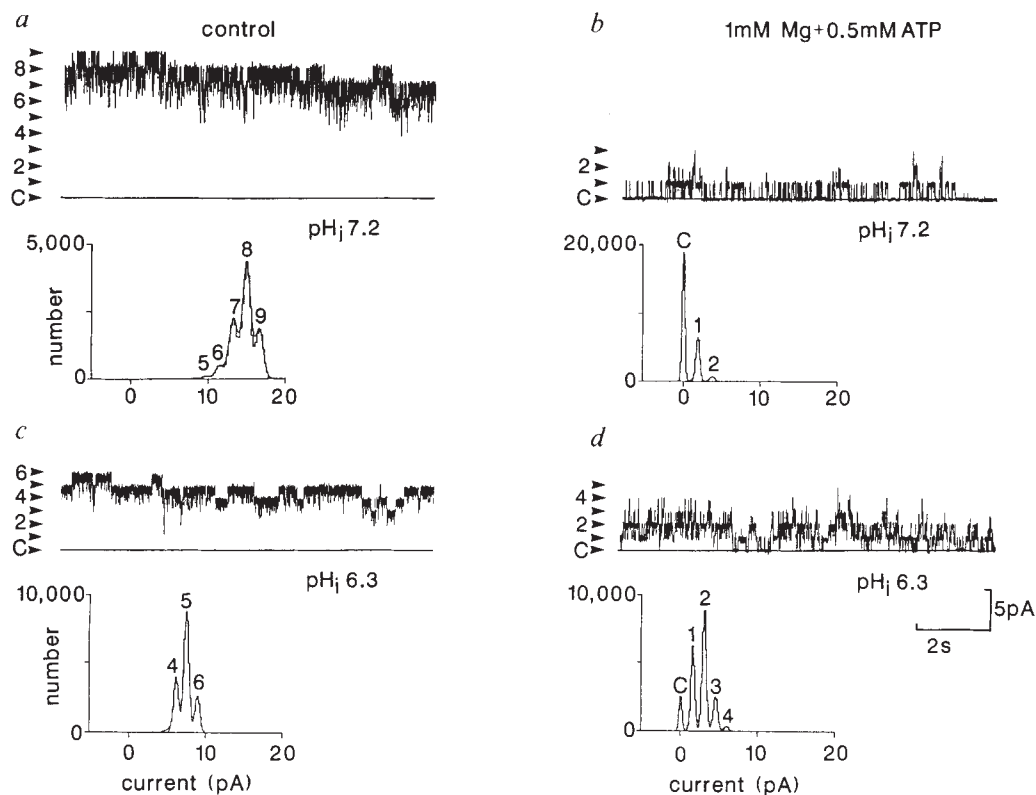
Although in the absence of ATP, P_{open} remained relatively constant with changing pH_i , there was a large reduction in the inhibitory effect of ATP as pH_i decreased. The fractional activity remaining after the application of 0.5 mM ATP was only 0.03 at pH_i 7.2, but ninefold greater (0.27) at pH_i 6.3 (means of three

FIG. 1 Effect of decreasing pH_i from 7.2 to 6.3 on the activity of a single K_{ATP} channel in an inside-out patch. The patch pipette contained 10 mM KCl, 110 mM NaCl, 1.8 mM $CaCl_2$ and 5 mM HEPES buffer, pH 7.2. The patch was placed in the outflow of a perfusion pipette²³, allowing exposure of the cytoplasmic side of the membrane to different solutions. These solutions contained KCl and KOH to bring the concentration of K^+ to 120 mM, 5 mM EGTA, 10 mM HEPES buffer for a pH of 8.0 and 7.2, 10 mM PIPES buffer for a pH of 6.3, and 10 mM MES buffer for a pH of 5.7. Currents were recorded with a List EPC-7 patch-clamp amplifier, stored on video tape and later filtered at 2 kHz and digitized at 10 kHz (CED 502-ADC) for analysis with a PDP11/73 minicomputer. The membrane potential was -3 mV so that the net current through the channel was outward ($E_K = -62$ mV). The recordings on the left illustrate that closings from the open state become less frequent with decreasing pH_i (C, closed; O, open). This effect is shown quantitatively by the open and closed time distributions to the right of each trace. These distributions were fitted by maximum-likelihood to a double-exponential probability density function of the form $f(t) = (a_1/\tau_1) \exp(-t/\tau_1) + (a_2/\tau_2) \exp(-t/\tau_2)$, where a_1 and a_2 denote the relative areas of component 1 and 2, t is time (in ms), and τ_1 and τ_2 are the time constants of component 1 and 2. For open times, τ_1 and τ_2 (in ms) are 1.18 and 6.31 for pH 7.2, and 1.21 and 13.88 for pH 6.3, and for closed times the values are 0.18 and 0.93 for pH 7.2, and 0.12 and 1.59 for pH 6.3. The mean of the open distributions (uncorrected for missed closings) are 5.77 and 12.27 ms for pH 7.2 and 6.3, respectively, and the mean of the



closed times (also uncorrected) are 0.32 and 0.24 ms for pH 7.2 and 6.3, respectively. These mean values are indicated by the arrows on the abscissa of each histogram. Because of the large number of brief closures, the measured mean open times were multiplied by the proportion of closed events detected to give the true mean open times²⁴. With a filter cutoff frequency of 2 kHz, the minimum-event duration detected by the half-level threshold is 90 μ s, so integration of the fitted closed-time distributions between 90 μ s and time infinity gives the proportion of detected closures. The corrected mean open times for pH 7.2 and 6.3 were 3.81 and 6.32 ms, respectively. Abbreviation: obs., observations.

FIG. 2 Effect of pH_i on the activity of K_{ATP} channels from an inside-out patch recorded in the absence and presence of Mg^{2+} and ATP. The pipette solution and control solutions at pH 7.2 and 6.3 are as described in Fig. 1. $MgCl_2$ (1 mM) and K_2 -ATP (0.5 mM) were added to the control solutions and the pH readjusted. In these experiments patches were exposed to each solution for 20–30 s. The records show segments of the activity of K_{ATP} channels in this patch in the absence and presence of Mg^{2+} and ATP at pH 7.2 (a, b) and at pH 6.3 (c, d). The inhibitory effect of Mg^{2+} and ATP was greater at pH 7.2 than at 6.3, P_{open} being 0.75, 0.04, 0.79 and 0.16 in a, b, c, and d, respectively. P_{open} was obtained by measuring the times, t_j , spent at the current levels corresponding to j channels open ($j=0, 1, 2, \dots, N$) and substituting these into $(\sum_{j=1}^N t_j j) / TN$, where T is the duration of the recording, and N the maximum number of simultaneously open channels seen under control conditions. Because of 'run-down' of channel activity, the number of active channels decreased from nine to six during the time between a and d, and was accounted for in the calculation of P_{open} . Similar results were obtained from other experiments in which the order of application of solutions was different. The amplitude histograms



below the traces were obtained from 7-s segments under each condition. The number above each peak corresponds to the current level at which 0(C), 1, 2, ... channels are open. Each peak was fitted with a gaussian distribution; the corresponding single-channel amplitudes were 1.90, 1.89, 1.48 and 1.50 pA in a, b, c and d, respectively.

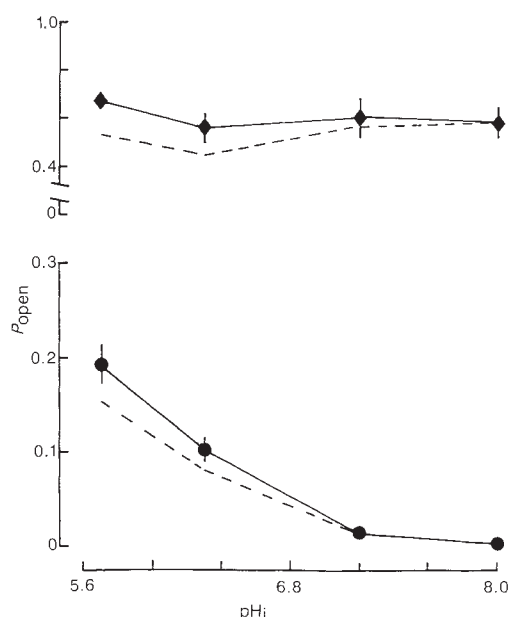


FIG. 3 The effect of pH_i on P_{open} in the absence (◆, upper graph) and presence (●, bottom) of 1 mM $MgCl_2$ and 0.5 mM ATP. Note the difference in ordinate scales. The data are pooled from 16 patches and points show the mean $\pm$ s.e.m., unless obscured by the symbol. In some patches there were too many channels to measure P_{open} , so values in the presence of Mg^{2+} and ATP (●) were calculated by multiplying the relative current in Mg^{2+} and ATP, normalized to the control current in the same patch, by the mean P_{open} (◆) measured from patches where this could be resolved under control conditions. The dashed lines show the effective ' P_{open} ' after accounting for the reduction in unitary amplitude, relative to that at pH_i 8.0, caused by a decrease in pH_i .

experiments). This change in blocking efficiency was also observed when the ATP was accompanied by excess Mg^{2+} ; under these conditions the predominant form of ATP is $Mg\text{-}ATP^{2-}$. The concentration of this complex is 0.36 mM at pH_i 7.2, and is reduced by only 14% to 0.31 mM at pH_i 6.3. In contrast to the block of these channels in pancreatic β -cells, there was relatively little difference between the block by 0.5 mM ATP in the absence and presence of Mg^{2+} , indicating that $Mg\text{-}ATP^{2-}$ is also effective at inhibiting K_{ATP} -channel activity in skeletal muscle. Figure 2 shows the results of an experiment with a patch held at -3 mV where the action of Mg^{2+} and ATP is compared with control conditions at pH_i 7.2 and 6.3. It is evident that the block by Mg^{2+} and ATP is much less at pH_i 6.3 than at 7.2. In five experiments the mean fractional activity remaining at pH_i 7.2 was 0.02 ± 0.01 , compared with 0.18 ± 0.02 at pH_i 6.3. In contrast to the inhibitory effect of Mg^{2+} on K_{ATP} -channel activity in cardiac cells¹⁷, 0.5 mM Mg^{2+} alone had little if any effect on the activity of these channels in skeletal muscle at a pH_i of either 8.0 or 5.7 (see also ref. 9).

The relationship between P_{open} and pH_i in the presence and absence of Mg^{2+} and ATP at a membrane potential of -3 mV is shown in Fig. 3. In the presence of Mg^{2+} and ATP, P_{open} increased markedly with decreasing pH_i , whereas in the absence of ATP, P_{open} remained virtually constant. This was also seen in patches held at other membrane potentials. In two patches at -33 mV, the mean fractional current remaining after application of 1 mM $MgCl_2$ and 0.5 mM ATP was 0.28 at pH_i 5.7, whereas the mean of four patches at pH_i 7.2 was only 0.01. In another patch held at $+27$ mV, the fractional current remaining at pH_i 5.7 was 0.52 compared with 0.04 at pH_i 8.0. The activity of K_{ATP} channels was therefore highly sensitive to changes in pH_i in the presence of Mg^{2+} and ATP at all the membrane potentials studied. The most sensitive region was between pH_i 7.2 and 6.0, which is within the physiological range for skeletal muscle^{13,14,18}. Almost all other types of K^+ currents, including inward and delayed rectifier currents, and Ca^{2+} -acti-

vated K^+ currents, are decreased by a fall in pH_i (refs 19–21). A decrease in pH_i has been shown to cause a reduction in the activity of K_{ATP} channels in pancreatic β -cells, although only in the presence of ATP²². But the K_{ATP} channels of skeletal muscle are unusual, as in the presence of Mg^{2+} and ATP, both of which are constituents of intracellular solution, a decrease in pH_i leads to an increase in the average current flowing through the channels. These results indicate that intracellular pH is important in regulating P_{open} of K_{ATP} channels in skeletal muscle. It will be interesting to see whether such an effect also occurs in cardiac and smooth muscle. □

Received 6 June; accepted 24 November 1989.

1. Noma, A. *Nature* **305**, 147–148 (1983).
2. Cook, D. L. & Hales, C. N. *Nature* **311**, 271–273 (1984).
3. Spruce, A. E., Standen, N. B. & Stanfield, P. R. *Nature* **316**, 736–738 (1985).
4. Standen, N. B. *et al.* *Science* **245**, 177–180 (1989).
5. Sturgess, N. C., Hales, C. N. & Ashford, M. L. T. *Pflügers Arch. ges. Physiol.* **409**, 607–165 (1987).
6. Noma, A. & Shibasaki, T. *J. Physiol., Lond.* **363**, 463–480 (1985).
7. Castle, N. A. & Haylett, D. G. *J. Physiol., Lond.* **383**, 31–43 (1987).
8. Carlson, F. D. & Siger, A. *J. gen. Physiol.* **44**, 33–60 (1960).
9. Spruce, A. E., Standen, N. B. & Stanfield, P. R. *J. Physiol., Lond.* **382**, 213–237 (1987).
10. Dunne, M. J. & Petersen, O. H. *FEBS Lett.* **208**, 58–62 (1986).
11. Dunne, M. J. & Petersen, O. H. *Pflügers Arch. ges. Physiol.* **407**, 564–565 (1986).
12. Kakei, M., Kelly, R. P., Ashcroft, S. J. H. & Ashcroft, F. M. *FEBS Lett.* **208**, 63–66 (1986).
13. Pan, J. W., Hamm, J. R., Rothman, D. L. & Shulman, R. G. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7836–7839 (1988).
14. Renaud, J. M. *J. Physiol., Lond.* **416**, 31–47 (1989).
15. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85–100 (1981).
16. Standen, N. B., Stanfield, P. R., Ward, T. A. & Wilson, S. W. *Proc. R. Soc.* **B221**, 455–464 (1984).
17. Findlay, I. *Pflügers Arch. ges. Physiol.* **410**, 313–320 (1987).
18. Challis, R. A. J., Vranic, M. & Radda, G. K., *Am. J. Physiol.* **256**, E129–E137 (1989).
19. Moody, W. J. *rev. Neurosci.* **7**, 257–278 (1984).
20. Cook, D. L., Masatoshi, I. & Fujimoto, W. Y. *Nature* **311**, 269–271 (1984).
21. Christensen, O. & Zeuthen, T. *Pflügers Arch. ges. Physiol.* **408**, 249–259 (1987).
22. Misler, S., Gillis, K. & Tabacharni, J. *J. Membrane Biol.* **109**, 135–143 (1989).
23. Carbone, E. & Lux, H. D. *J. Physiol., Lond.* **386**, 547–570 (1987).
24. Colquhoun, D. & Sigworth, F. J. in *Single Channel Recording* (eds Neher, E. & Sakmann, B.) 191–263 (Plenum, New York, 1983).

ACKNOWLEDGEMENTS. I am grateful to Professor P. R. Stanfield and Dr N. B. Standen for their constructive criticism during the preparation of this manuscript, and the MRC for financial support.

Phosphorylation of GAP and GAP-associated proteins by transforming and mitogenic tyrosine kinases

Christine Ellis, Michael Moran, Frank McCormick* & Tony Pawson†

Division of Molecular and Developmental Biology, Mount Sinai Hospital Research Institute, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada

* Department of Molecular Biology, Cetus Corporation, Emeryville, California 94608, USA

THE critical pathways through which protein-tyrosine kinases induce cellular proliferation and malignant transformation are not well defined. As microinjection of antibodies against $p21^{ras}$ can block the biological effects of both normal and oncogenic tyrosine kinases, it is likely that they require functional $p21^{ras}$ to transmit their mitogenic signals^{1,2}. No biochemical link has been established, however, between tyrosine kinases and $p21^{ras}$. We have identified a non-catalytic domain of cytoplasmic tyrosine kinases, SH2, that regulates the activity and specificity of the kinase domain^{3–9}. The presence of two adjacent SH2 domains in the $p21^{ras}$ GTPase-activating protein (GAP)^{6,10,11} indicates that GAP might interact directly with tyrosine kinases. Here we show that GAP, and two co-precipitating proteins of relative molecular masses 62,000 and 190,000 (p62 and p190) are phosphorylated on tyrosine

† To whom correspondence should be addressed.

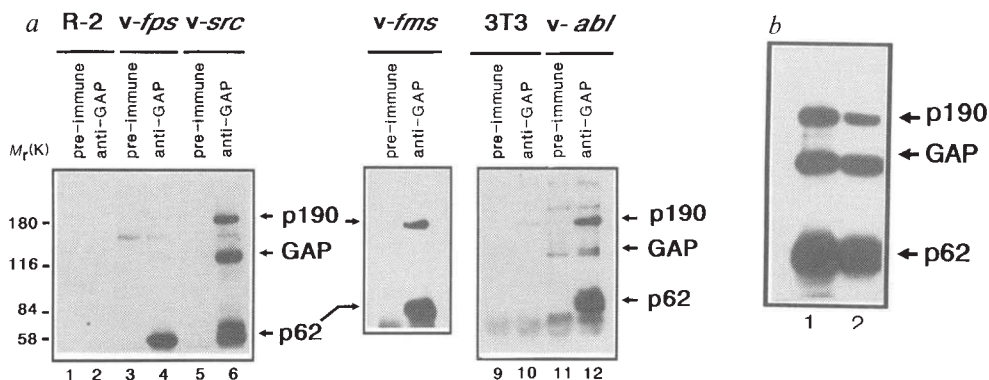
in cells that have been transformed by cytoplasmic and receptor-like tyrosine kinases. The phosphorylation of these polypeptides correlates with transformation in cells expressing inducible forms of the *v-src* or *v-fps* encoded tyrosine kinases. Furthermore, GAP, p62 and p190 are also rapidly phosphorylated on tyrosine in fibroblasts stimulated with epidermal growth factor. Our results suggest a mechanism by which tyrosine kinases might modify p21^{ras} function, and implicate GAP and its associated proteins as targets of both oncoproteins and normal growth factor receptors with tyrosine kinase activity. These data support the idea⁴⁻⁶ that SH2 sequences direct the interactions of cytoplasmic proteins involved in signal transduction.

Rabbit antisera were raised against a trpE-GAP bacterial fusion protein containing amino-acid residues 171-448 from the N-terminal region of human GAP (ref. 11). Rat or mouse fibroblast cell lines transformed by the *v-fps*, *v-src* or *v-abl* retroviral oncogenes, which encode membrane-associated cytoplasmic tyrosine kinases (P130^{gag-fps}, p60^{v-src} and P120^{gag-abl} respec-

tively), were lysed and immunoprecipitated with anti-GAP(171-448) antiserum. Western blotting of the precipitated proteins with affinity-purified anti-phosphotyrosine antibodies identified three polypeptides from the transformed cells, with mobilities corresponding to M_r s of 62,000 (62K), 124K and 190K (Fig. 1a). The 62K band was most prominent in each case, and migrated as a doublet in *v-src* and *v-abl*-transformed cells. Although the same phosphotyrosine-containing proteins were present in anti-GAP(171-448) immunoprecipitates from the different transformed cells, their relative abundance varied. In particular, tyrosine phosphorylation of the 124K protein, subsequently shown to be GAP, was highest in *v-src*-transformed Rat-2 cells (Rat-2 *v-src*). Immunoprecipitation of Rat-2 *v-src* cells with anti-peptide antibodies directed against residues 975-990 of human GAP yielded a pattern of phosphotyrosine-containing proteins identical to that observed with anti-GAP(171-448) antiserum (fig. 1b). Phosphotyrosine-containing proteins precipitated by anti-GAP antiserum were specific to

FIG. 1 GAP, p62 and p190 from cells transformed by tyrosine kinase oncoproteins are recognized by anti-phosphotyrosine antibodies. **a**, Immunoprecipitates from cell lysates using pre-immune serum (lanes 1, 3, 5, 7, 9, 11) or anti-GAP(171-448) antiserum (lanes 2, 4, 6, 8, 10, 12), were separated by electrophoresis on a 7.5% SDS-polyacrylamide gel, transferred to nitrocellulose, and blotted with affinity-purified rabbit anti-phosphotyrosine antibodies followed by ¹²⁵I-protein A. Immunoprecipitates were from normal Rat-2 cells (lanes 1 and 2), Rat-2 cells transformed by avian *v-fps* (lanes 3 and 4), avian *v-src* (lanes 5 and 6) or feline *v-fms* (lanes 7 and 8), NIH 3T3 cells (lanes 9 and 10) or *v-abl*-transformed NIH 3T3 cells (lanes 11 and 12). **b**, Rat-2 *v-src* cells were immunoprecipitated with (lane 1) anti-GAP(171-448) or (lane 2) anti-GAP(975-990) antibodies, and immunoblotted with anti-phosphotyrosine antibodies.

METHODS. To obtain anti-GAP(171-448) antiserum, a *BalI*-*SalI* restriction fragment from a human GAP complementary DNA was subcloned into the pATH11 bacterial expression vector. The 66 K trpE-GAP protein was induced, purified, and used to immunize rabbits²⁰. Proteins were immunoprecipitated as detailed elsewhere¹⁸. Briefly, cells were lysed on ice in 1 ml 50 mM HEPES, pH 7.5, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1.5 mM MgCl₂, 1 mM EGTA, 10 µg ml⁻¹ aprotinin, 10 µg ml⁻¹ leupeptin, 1 mM PMSF, 200 µM sodium orthovanadate, 10 mM pyrophosphate and 100 mM sodium fluoride.



As minor fractions of the highly phosphorylated Gag-containing tyrosine kinases tend to precipitate non-specifically under these lysis conditions, pre-immune serum was always included as a negative control. About 2×10^6 cells were used in each immunoprecipitation, except for samples run in lanes 7 and 8, where $\sim 1 \times 10^7$ cells were used. Lysates were centrifuged for 30 min at 10,000g and the supernatants incubated for 90 min at 4 °C with anti-GAP antiserum or pre-immune serum and 100 µl 20% protein A-Sepharose. The resulting immune complexes were washed three times with 20 mM HEPES, pH 7.5, 10% glycerol, 0.1% Triton X-100, 150 mM NaCl and 1 mM sodium orthovanadate, heated in SDS sample buffer, separated by gel electrophoresis, and immunoblotted with anti-phosphotyrosine antibodies^{6,21}. The isolation and specificity of these antibodies has been described^{6,21,22}.

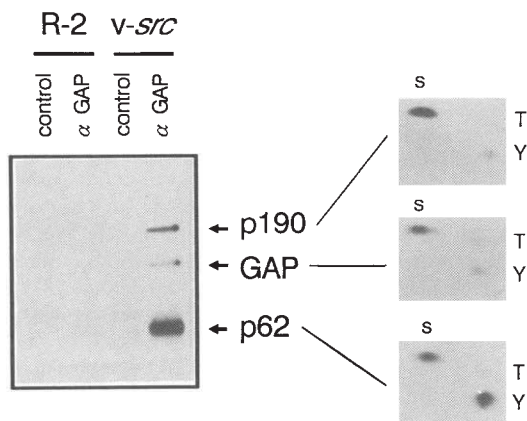


FIG. 2 Phosphorylation of GAP, p62 and p190 is induced by *v-src* transformation. Rat-2 cells and Rat-2 *v-src* cells were metabolically labelled with ³²P_i, immunoprecipitated with control rabbit antibodies or affinity-purified anti-GAP(171-448) antibodies, separated by gel electrophoresis, transferred to an Immobilon membrane and autoradiographed (on the left). Phosphorylated GAP, p62 and p190 from Rat-2 *v-src* cells were analysed for phosphoamino acids (on the right). S, phosphoserine; T, phosphothreonine; Y, phosphotyrosine.

METHODS. 5×10^6 Rat-2 cells or Rat-2 *v-src* cells in 10-cm plates were incubated for 2 h in phosphate-free Dulbecco's modified Eagle medium containing 2% dialysed fetal bovine serum and 5 mCi ³²P_i. Lysates were prepared as described in the legend to Fig. 1, and were precleared once by incubation for 1 h with protein-A agarose beads coated with normal rabbit serum, and again by incubation with protein-A agarose beads coated with 10 µg affinity-purified rabbit anti-mouse immunoglobulin antibodies. The precleared lysates were divided into two and incubated with protein-A agarose beads coated with either 2 µg rabbit anti-mouse immunoglobulin antibodies (as a control), or with 2 µg affinity-purified rabbit anti-GAP(171-448) antibodies. The immunoprecipitates were washed and resolved by gel electrophoresis as for Fig. 1, electrophoretically transferred to an Immobilon membrane (Millipore) and exposed to XAR film at -70 °C for 10 h with an intensifying screen. Regions of the membrane containing the indicated ³²P-labelled proteins were incubated with 5.7 N HCl at 110 °C for 1 h²³. The resulting hydrolysates were analysed by two-dimensional thin-layer electrophoresis²⁴ and exposed to film for 85 h with an intensifying screen.

the transformed cells, with the exception of p190, which was detectable at low levels in parental fibroblasts (Fig. 1a).

Metabolic labelling of normal or *v-src*-transformed Rat-2 cells with $^{32}\text{P}_i$ followed by immunoprecipitation with anti-GAP antibodies, showed that the total phosphorylation of GAP, p62 and p190 was strongly induced in the Rat-2 *v-src* cells (Fig. 2). Phosphoamino-acid analysis indicated that 65% of the phosphate in p62 from Rat-2 *v-src* cells was in the form of phosphotyrosine, compared with 14% for GAP and 7% for p190 (Fig. 2). Therefore p62 is highly phosphorylated on tyrosine in transformed cells, whereas the stoichiometry of GAP tyrosine phosphorylation seems to be low (Fig. 2; and unpublished results). But, the molecules of GAP containing phosphotyrosine could represent a functionally important subpopulation. All three pro-

teins from transformed cells contained both phosphoserine and phosphothreonine, in addition to phosphotyrosine, and indeed phosphoserine was the chief phosphoamino acid of GAP and p190.

The phosphotyrosine-containing proteins precipitated by anti-GAP(171-448) antiserum were blotted with the same antiserum, or with antibodies to the GAP(975-990) peptide (Fig. 3). Both anti-GAP antibodies recognized only a single main band, the mobility of which was identical to that of 124K protein previously identified with anti-phosphotyrosine antibodies in transformed cells (compare Fig. 1a, lanes 3-6, with Fig. 3a). These results confirm that the 124K protein phosphorylated on tyrosine in cells transformed by *v-fps*, *v-src* or *v-abl* is GAP itself. The expression of GAP, in contrast to its phosphorylation, was equivalent in both normal and transformed cells (Fig. 3b). Potentially, p62 and p190 could be precipitated because they are complexed with GAP, or could be recognized directly by anti-GAP antibodies. As p62 and p190 were precipitated by antibodies both to the N- and the C-terminal regions of GAP, but were not recognized by either antibody on immunoblots, direct antibody binding seems unlikely. In anti-GAP immunoprecipitates of [^{35}S]methionine-labelled *v-src*-transformed cells, p190 was readily identified but it was not found in immunoprecipitates of normal Rat-2 cells. Boiling lysates of transformed cells in 0.5% SDS abolished the subsequent immunoprecipitation of p190 and p62, but not of GAP (data not shown). We conclude that p190 and p62 form a specific complex with GAP. As p62 did not seem to bind either antibody on a western blot, it is unlikely to be a proteolytic fragment of GAP, and indeed it has a distinct tryptic phosphopeptide map (data not shown).

To extend these observations to receptor-like tyrosine kinases, we first examined Rat-2 cells transformed by *v-fms*, which encodes an activated form of the macrophage colony stimulating factor receptor¹² (Fig. 1a). Anti-GAP immunoprecipitates from these cells contained tyrosine-phosphorylated p62 and p190, although about fivefold more cells were required to generate a signal equivalent to that observed with the cytoplasmic oncoproteins. GAP from *v-fms*-transformed cells was poorly reactive with anti-phosphotyrosine antibodies, but could be clearly detected after prolonged exposure. These observations suggest that transmembrane as well as cytoplasmic tyrosine kinases interact directly with GAP. To investigate GAP phosphorylation after growth factor stimulation we used Rat-1 fibroblasts expressing $\sim 2.5 \times 10^5$ human epidermal growth factor (EGF)-receptors per cell (W. Wasilenko and M. Weber, personal communication). EGF-treatment of starved cells induced phosphorylation of p62, GAP and p190 within 2.5 min (Fig. 4a), suggesting that they are substrates for the EGF-receptor. There was a similar rapid phosphorylation of GAP and GAP-associated proteins after stimulation of Rat-2 cells with platelet-derived growth factor (data not shown). Phosphorylation of GAP, p62 and p190 is therefore a feature of the normal signal transduction process in fibroblasts.

We used two cell lines that are inducible for *v-fps* or *v-src* transforming activity to investigate the relationship between the phosphorylation of GAP and GAP-associated proteins and transformation. The addition of heavy metals to a Rat-2 cell line that contains the coding sequence for P130^{gag-fps} under the control of the human metallothionein-IIA promoter, induces a 14-fold increase in P130^{gag-fps} expression over the course of 24 h, resulting in their conversion from a normal to a transformed phenotype (M.M., unpublished results). Untreated cells, or cells collected after 24 h of heavy-metal treatment, were immunoprecipitated with an anti-Gag monoclonal antibody (which specifically recognizes P130^{gag-fps}) or with anti-GAP antiserum, and the immunoprecipitates were blotted with anti-phosphotyrosine antibodies (Fig. 4b). The basal level of tyrosine phosphorylation is higher in these cells than in normal Rat-2 cells owing to residual *v-fps* expression in the absence of zinc

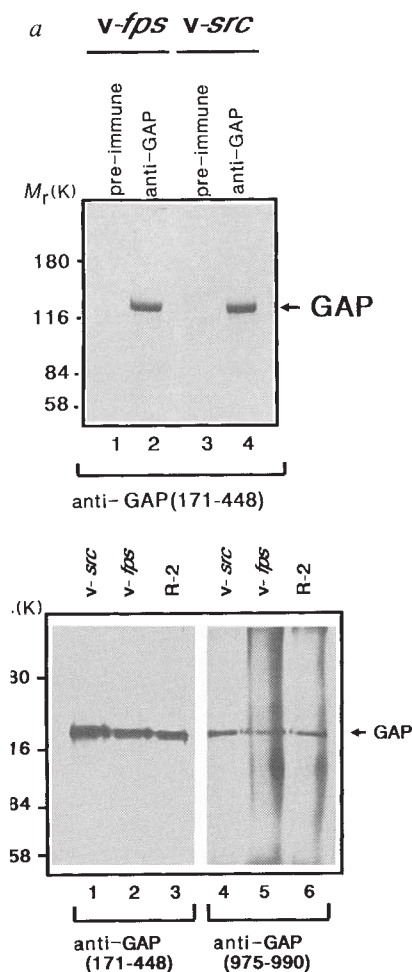


FIG. 3 The 124 K protein recognized by anti-phosphotyrosine antibodies is GAP. *a*, Immunoprecipitates obtained with pre-immune serum (lanes 1 and 3) or anti-GAP(171-448) antiserum (lanes 2 and 4) from Rat-2 cells transformed by *v-fps* (lanes 1 and 2) or *v-src* (lanes 3 and 4) were first blotted with anti-phosphotyrosine antibodies (Fig. 1a, lanes 3-6). The nitrocellulose was then stripped of antibody and reprobed with anti-GAP(171-448). *b*, Rat-2 cells transformed with *v-src* (lanes 1 and 4) or *v-fps* (lanes 2 and 5), or parental Rat-2 cells (lanes 3 and 6) were lysed and immunoprecipitated with anti-GAP(171-448) antiserum. Precipitates were immunoblotted directly with anti-GAP(171-448) antiserum (lanes 1-3) or affinity-purified anti-GAP(975-990) antibodies (lanes 4-6). Migration of molecular weight calibration standards is shown on the left ($\times 10^{-3}$).

METHODS. In *a*, anti-phosphotyrosine antibodies were stripped from the western blot by two successive 15 min incubations with a solution containing 5 M sodium iodide and 1 mM sodium thiosulphate, followed by a brief rinse in distilled water. After autoradiography to confirm that all radioactivity was lost, the stripped blot was reprobed with anti-GAP(171-448) antiserum. In both *a* and *b* bound antibody was identified using an alkaline phosphatase-conjugated secondary antibody, followed by incubation with nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate⁶.

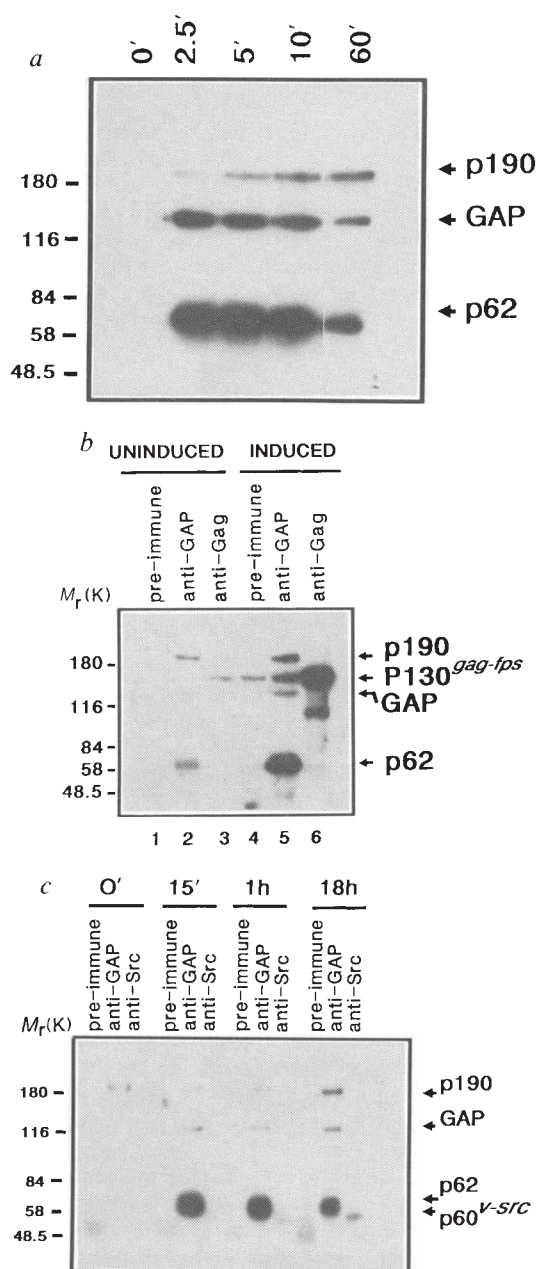


FIG. 4 Tyrosine phosphorylation of GAP and GAP-associated proteins in response to EGF stimulation, and induction of *v-fps* or *v-src* kinases. **a**, Rat-1 cells expressing 2.5×10^5 human EGF-receptors were grown to confluence and serum-starved for 48 h in medium containing 0.5% fetal bovine serum. Starved cells were incubated with 80 nM EGF at 37 °C for 0–60 min as indicated. Cells were lysed, immunoprecipitated with anti-GAP(171–448) antibodies and immunoblotted with anti-phosphotyrosine antibodies. **b**, Rat-2 cells containing the coding sequence for P130^{gag-fps} under the control of the metallothionein-IIA promoter were grown in the absence (uninduced; lanes 1–3) or presence (induced; lanes 4–6) of 10 μ M CdCl₂ and 50 μ M ZnSO₄ for 24 h. Cells were lysed and immunoprecipitated with pre-immune serum (lanes 1 and 4), with anti-GAP(171–448) antiserum (lanes 2 and 5), or with an anti-Gag monoclonal antibody (lanes 3 and 6). Precipitates were immunoblotted with anti-phosphotyrosine antibodies. **c**, Rat-1 tsLA29 *src* cells, expressing a p60^{v-src} temperature sensitive for tyrosine kinase and transforming activities²⁵, were incubated at the non-permissive temperature of 39.5 °C for 18 h, and then maintained at 39.5 °C, or shifted to the permissive temperature of 35 °C for 15 min, 1 h or 18 h. Cells were lysed and immunoprecipitated with pre-immune serum, anti-GAP(171–448) antiserum or anti Src monoclonal antibody 327. Precipitates were immunoblotted with anti-phosphotyrosine antibodies. Immunoprecipitations and anti-phosphotyrosine antibody blotting were as described in the legend to Fig. 1. Migration positions of molecular weight standards ($\times 10^{-3}$) are shown on the left for **a**, **b** and **c**.

and cadmium. Coordinately with the increased expression of P130^{gag-fps}, anti-GAP immunoprecipitates showed an increase in the tyrosine phosphorylation of p62, GAP and p190. Similar results were obtained with Rat-1 cells expressing a mutant p60^{v-src} (tsLA29 *src*) which is temperature-sensitive for kinase and transforming activities (Fig. 4c). Fifteen minutes after shifting Rat-1 tsLA29 cells from the non-permissive to the permissive temperature, the tyrosine phosphorylation of p62 and GAP was markedly enhanced, indicating that they are substrates of p60^{v-src} (Fig. 4c). There was also a delayed and minor increase in p190-tyrosine phosphorylation. The induction of p62 phosphorylation in all systems was particularly rapid and striking.

These results have several interesting implications. Previous work has indicated that functional p21^{ras} is required for the oncogenic and mitogenic activities of tyrosine kinases^{1,2}. The tyrosine phosphorylation of GAP forges a direct link between tyrosine kinases and *ras*-encoded proteins, because GAP in turn regulates p21^{ras} GTPase activity¹³. The apparent association of p62 and p190 with GAP indicates the presence of a heteromeric complex that may couple p21^{ras} to upstream regulators (such as tyrosine kinases) and downstream targets. The identities of the 62K and 190K proteins are not known. Interestingly, cross-linking experiments have identified a 60K protein closely associated with p21^{ras} (ref. 14). Although we have no immediate evidence that phosphorylation of GAP or GAP-associated proteins modulates p21^{ras} function, regulatory steps such as p21^{ras} guanine nucleotide exchange, stimulation of p21^{ras} GTPase activity, or the formation of an activated p21^{ras}/GAP complex at the membrane might be modified by tyrosine phosphorylation of one or all of these proteins.

Are GAP and its associated proteins biologically important substrates of oncogenic tyrosine kinases? Our data are consistent with this hypothesis, although they do not prove it. GAP, p62 and p190 are phosphorylated in cells transformed by a variety of tyrosine kinases, as might be expected of proteins whose phosphorylation is critical for transformation. A comparative analysis of the phosphorylation of GAP and GAP-associated proteins in cells expressing wild type and transformation-defective tyrosine kinase oncoproteins will be necessary to resolve this issue.

How could tyrosine kinases and GAP interact with one another? GAP contains two copies of the SH2 domain, originally identified as a regulatory element of cytoplasmic tyrosine kinases³, and subsequently found in phospholipase C_γ and the *v-crk* oncoprotein^{15,16}. A probable function of the conserved SH2 domains of these latter proteins is to direct intermolecular interactions with tyrosine kinases. Thus phospholipase C_γ is a substrate for some growth factor receptors^{17–19}, whereas P47^{gag-crk} activates an endogenous cellular tyrosine kinase¹⁶. The finding that GAP and GAP-associated proteins are phosphorylated on tyrosine in EGF-stimulated and *v-src*-transformed cells lends credence to this hypothesis. □

Received 28 September; accepted 20 November 1989.

- Mulcahy, L. S., Smith, M. R. & Stacey, D. W. *Nature* **313**, 241–243 (1985).
- Smith, M. R., DeGudicibus, S. J. & Stacey, D. W. *Nature* **320**, 540–543 (1986).
- Sadowski, I., Stone, J. C. & Pawson, T. *Molec. cell. Biol.* **6**, 4396–4408 (1986).
- DeClue, J. E., Sadowski, I., Martin, G. S. & Pawson, T. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9064–9068 (1987).
- Pawson, T. *Oncogene* **3**, 491–495 (1988).
- Koch, C. A., Moran, M., Sadowski, I. & Pawson, T. *Molec. cell. Biol.* **10**, 4131–4140 (1989).
- Raymond, V. W. & Parsons, J. T. *Virology* **160**, 400–410 (1987).
- Wang, H.-C. R. & Parsons, J. T. *J. Virol.* **63**, 291–302 (1989).
- Verderame, M. F., Kaplan, J. M. & Varmus, H. E. *J. Virol.* **63**, 338–348 (1989).
- Vogel, V. S. *et al.* *Nature* **335**, 90–93 (1988).
- Trahey, M. *et al.* *Science* **242**, 1697–1700 (1988).
- Sherr, C. J. *et al.* *Cell* **41**, 665–676 (1985).
- Trahey, M. & McCormick, F. *Science* **238**, 542–545 (1987).
- De Gunzburg, J. D., Riehl, R. & Weinberg, R. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4007–4011 (1989).
- Stahl, M. L., Ferenz, C. R., Kelleher, K. L., Kriz, R. W. & Knopf, J. *Nature* **332**, 269–272 (1988).
- Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
- Wahl, M. L., Nishibe, S., Suh, P.-G. & Carpenter, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1568–1572 (1989).
- Margolis, B. *et al.* *Cell* **57**, 1101–1107 (1989).
- Meisenhelder, J., Suh, P.-G., Rhee, S. G. & Hunter, T. *Cell* **57**, 1109–1122 (1989).
- Sadowski, I. & Pawson, T. *Oncogene* **1**, 181–191 (1987).

21. Kamps, M. P. & Sefton, B. M. *Oncogene* **2**, 305–315 (1988).
22. Letwin, K., Yee, S.-P. & Pawson, T. *Oncogene* **3**, 621–627 (1988).
23. Kamps, M. P. & Sefton, B. M. *Analyt. Biochem.* **176**, 22–27 (1989).
24. Cooper, J. A., Sefton, B. M. & Hunter, T. *Meth. Enzym.* **99**, 387–402 (1983).
25. Johnson, R. M., Wasilenko, W. J., Mattingly, R. R., Weber, M. J. & Garrison, J. C. *Science* **246**, 121–124 (1989).

ACKNOWLEDGEMENTS. We thank Theresa Lee and Laurie Gray for technical assistance, Peter Greer and Gareth Taylor for advice and help with bacterial expression and immunization, Kenneth Letwin for anti-phosphotyrosine antibodies and Anne Koch for cell lines. We are indebted to Michal Weber for Rat-1 cell lines expressing the human EGF-receptor and tsLA29 src, and to Jeff Kudlow for EGF. This work was supported by the National Cancer Institute of Canada and the Medical Research Council of Canada.

Immunological activity of covalently linked T-cell epitopes

Francesco Ria^{*†}, Bosco M. C. Chan^{*}, Mark T. Scherer^{*}, John A. Smith[‡] & Malcolm L. Gefter^{*}

^{*} Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

[‡] Departments of Molecular Biology and Pathology, Massachusetts General Hospital and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02114, USA

IMMUNE responses to proteins necessarily involve the recognition by T lymphocytes of a peptide or peptides derived from a protein complexed with a major histocompatibility antigen. The T-cell response of BALB/c mice to the bacteriophage λ cI repressor protein (residues 1–102) is directed predominantly towards the epitope contained within a single peptide encompassing residues 12–26 (refs 1, 2). Similar phenomena of immunodominance of a particular peptide have also been observed in other protein systems^{3–6}. The mechanisms that have been suggested to account for the focusing of the T-cell response⁷ are partial deletion in the T-cell repertoire, biased antigen processing, and competition for binding to the presenting molecule, the major histocompatibility complex encoded class II transplantation antigen. In a model system with a polypeptide containing two synthetically linked immunologically active epitopes, we now demonstrate the existence of a hierarchy between these epitopes, so that the immune response elicited is directed mainly towards the more immunogenic epitope, whereas the less immunogenic epitope elicits little or no T-cell reactivity. In addition, the same hierarchy of dominance is also apparent when the polypeptide is used to induce tolerance in the periphery in adult mice. The chimaeric peptide can induce tolerance only towards the more immunogenic epitope. These experiments indicate that the rules governing antigen processing and presentation that result in T-cell activation are apparently the same as the rules that govern the processes resulting in the induction of tolerance.

We have synthesized a peptide containing residues 12–26 of the λ repressor protein cI (p12–26), joined through a Gly-Pro-Gly sequence to residues 325–336 of chicken ovalbumin (p325–336). This joined peptide contains the immunodominant epitopes of the two proteins expressed in BALB/c mice^{1,6}. The ovalbumin-derived peptide was modified at position 327 by the substitution of valine by aspartic acid (pOVAD). In confirmation of previous results⁸, this single amino-acid substitution decreased the apparent affinity of the p325–336 peptide for the I-A^d-encoded class II major histocompatibility complex (MHC) molecule (Fig. 1d), whereas the epitope recognized by the T-cell receptor within this peptide was apparently unaltered. The two distinct T-cell epitopes, p12–26 and pOVAD, in the joined peptide were each immunologically active *in vitro* (Fig. 1a–c). The joined peptide was able to stimulate both the λ repressor cI p12–26-specific T-cell hybridomas 9C127 and 1E1, and the ovalbumin-specific hybridoma 3D054.8. Depending on the

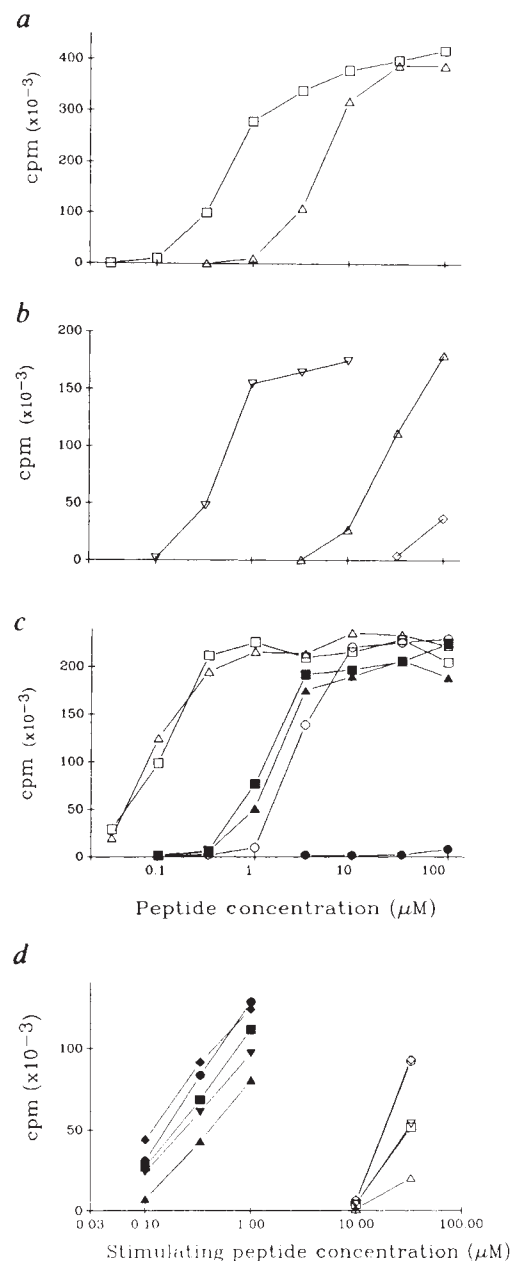


FIG. 1 a, b and c, *In vitro* activity of the joined peptide (p12–26-Gly-Pro-Gly-pOVAD). The ability of the joined peptide to stimulate λ repressor cI (12–26) and ovalbumin-specific hybridomas was tested using the hybridomas 9C127 (a), eD054.8 (b) and 1E1 (c). The hybridomas were stimulated with the joined peptide (Δ), p12–26 (□), pOVAD (◇), pOVAD(324–336) (▽), p1–102 of λ repressor protein cI (○). As antigen-presenting cells, L cells (a, b) and A20 (open symbols) or fixed A20 (closed symbols; c) were used. d, Ability of pOVAD (open symbols) and pOVA (324–336; closed symbols) to stimulate the 3D054.8 hybridoma when presented alone (●) or in presence of different concentrations of p12–26 (100 μM (▲), 50 μM (■), 25 μM (▼), 12.5 μM (◆)). A20 cells were used as antigen-presenting cells.

METHODS. Details of the preparation of the p12–26-specific hybridomas are described elsewhere¹⁶. The hybridoma 3D054.8 was kindly provided by Drs J. Kappler and P. Marrack. The L cells RT 2.3.3.H, expressing I-A^d-encoded molecule, were a gift of Dr R. N. Germain. The complete sequence of λ repressor protein is reported by Sauer and Andregg¹⁷. All peptides were synthesized by the solid phase method of Merifield¹⁸ using an automated Applied Biosystem 430A peptide synthesizer. The hybridomas (5×10^4 cells per well) were stimulated by APC cells (5×10^4 cells per well) in the presence of varying concentrations of test peptides. After 24 h, interleukin-2 secretion was measured previously described¹⁶. APC cells (A20) were fixed with 0.1% glutaraldehyde in PBS, and the reaction was quenched with 0.2 M lysine in PBS after 30 s.

[†] Permanent address: Istituto di Patologia Generale, Università Cattolica, L.go F. Vito 1, 00168 Rome, Italy.

hybridoma, the stimulatory activity of the joined peptide could differ from that observed when using either p12-26 or pOVAD alone. In addition, the 12-26 epitope in the joined peptide was able to be presented by antigen-presenting cells (APC) previously fixed with glutaraldehyde, indicating that processing of the joined peptide is not essential *in vitro* for peptide presentation.

We also investigated the immune response to the two distinct epitopes of the joined peptide *in vivo*. As a positive control we showed that the immunization with the pOVAD peptide induced T cells that were capable of being stimulated either by the immunogen, pOVAD, or by the joined peptide (Fig. 2a), as was seen *in vitro* with the ovalbumin-specific hybridoma. By contrast, mice immunized with the joined peptide (Fig. 2b) showed little response, if any, to the pOVAD peptide. But the same joined peptide did induce a response to p12-26. Also, the response of these T cells was generally greater when stimulated by the joined peptide than it was when stimulated by p12-26. This could be attributed either to generation of a junctional epitope or epitopes, or to enhancement of the response to p12-26 by the flanking amino acids in the joined peptide⁹. In either case, the response directed towards p12-26 was not reduced (Fig. 2b, f). But these results show that a hierarchy is established between the two distinct epitopes in the joined peptide in such a way

that the phenomenon of immunodominance, as observed in earlier studies¹⁻⁶, is demonstrated in this model system. The negative influence of p12-26 observed when it is covalently linked to pOVAD is probably due to competition for binding to the MHC-encoded molecules. Immunization with equal amounts of both peptides simultaneously resulted only in a slight inhibition of the response to the pOVAD epitope, compared with the response with pOVAD alone (Fig. 2c). That the response is lowered is consistent with *in vitro*¹⁰ and *in vivo*¹¹ results showing that immunoactive peptides compete for presentation. Immunization with a joined peptide in which p12-26 is covalently linked through the same Gly-Pro-Gly sequence to a wild-type ovalbumin residues 325-336 (pOVA), containing the original Val in position 327 (that is, a more immunologically active epitope) induces comparable, but already diminished, responses for both the 12-26 and the pOVA epitopes thus further supporting the existence of competition for responses between different epitopes *in vivo* (Fig. 2e).

The p12-26 epitope can be viewed as a suppressor epitope for the response to pOVAD. The effect of the presence of more than one potential epitope on the activity of another in this system can be related to the intrinsic binding characteristics of class II molecules, the processing mechanisms that can favour one epitope over another, or the presence of additional amino-

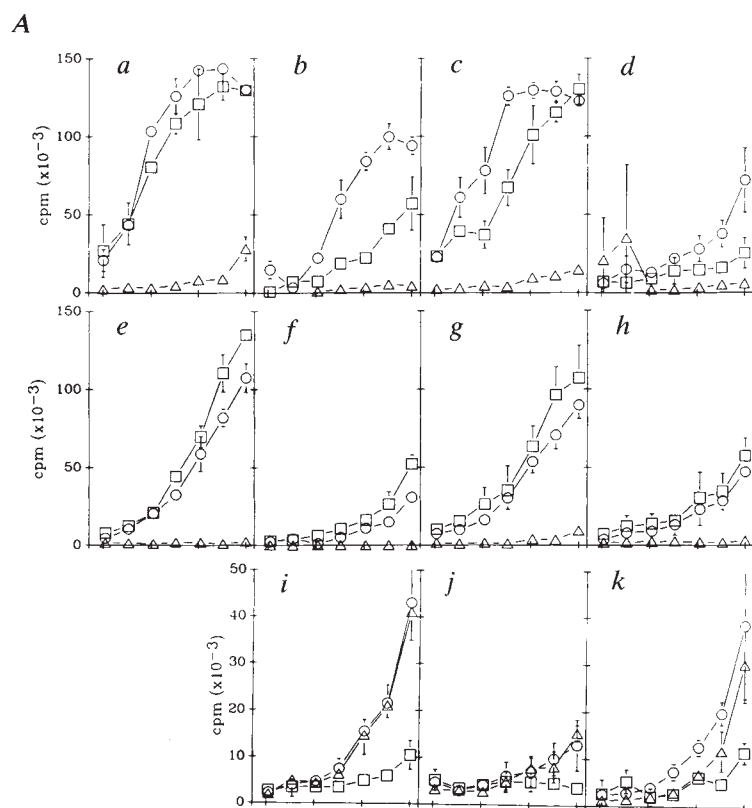
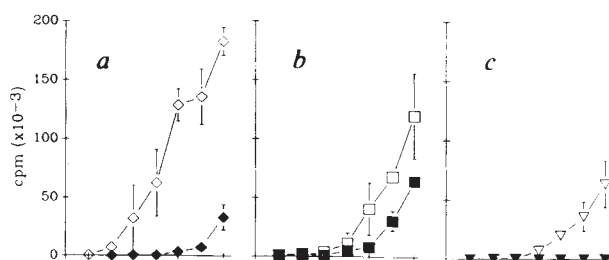


FIG. 3 A, Interleukin-2 production from lymphocytes of BALB/c mice immunized with the joined peptide p12-26-pOVAD (a, b, c, d), p12-26 (e, f, g, h) or pOVAD (i, j, k) after treatment with saline (a, e) or tolerization with p12-26 (b, f, i), pOVAD (c, g, j) or joined peptide p12-26-Gly-Pro-Gly-pOVA (d, h, k). The response was analysed by stimulating the lymphocytes with the same joined peptide, p12-26 and pOVAD. B, Interleukin-2 production from lymphocytes of BALB/c mice treated with saline (open symbols) or tolerized with the control joined peptide p12-26-Gly-Pro-Gly-pOVA (closed symbols) after immunization with the same joined peptide (a), p12-26 (b) and pOVA (325-336; c). Symbols and concentrations are the same as for Fig. 2.

METHODS. Adult BALB/c mice (6-10 weeks of age) were tolerized by intraperitoneal injection of 300 μ g peptide emulsified in incomplete Freund's adjuvant¹². Ten days later, mice from each group of the tolerized group were immunized with 100 μ g p12-26, pOVAD or pOVA (324-336), or 200 μ g joined peptides emulsified in CFA, subcutaneously. Interleukin-2 secretion of lymphocytes from the draining lymph nodes was measured as described elsewhere¹⁶. Four mice in each group were used, except for in the experiments of Af in which one mouse was not tolerized.

B



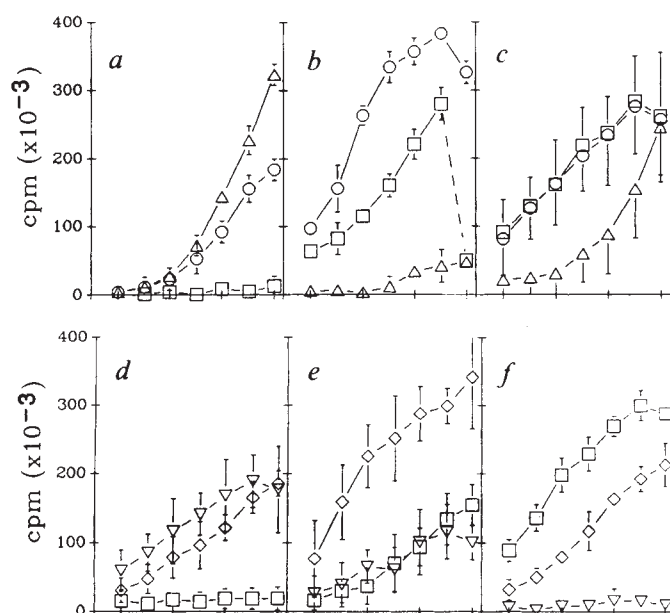


FIG. 2 Proliferation of lymphocytes from BALB/c mice immunized with pOVAD (a) or the joined peptide p12-26-Gly-Pro-Gly-pOVAD (b), pOVAD and p12-26 (c), pOVA (324-336; d), the control joined peptide p12-26-Gly-Pro-Gly-pOVA (e) and p12-26 (f) in response to stimulation with the joined peptide p12-26-Gly-Pro-Gly-pOVA (○), p12-26 (□), pOVA (Δ), pOVA (324-336; ▽), the joined peptide p12-26-Gly-Pro-Gly-pOVA (◇), at concentrations 100–0.1 μM.

METHODS. Four Balb/c mice were immunized with 100 μg pOVAD or p12-26, or with 200 μg joined peptides or 1:1 mixture of p12-26 and pOVAD emulsified in complete Freund's adjuvant, subcutaneously. Cells from the draining lymph nodes were collected, 7 days later, and proliferation and interleukin-2 secretion (data not shown) after stimulation with the peptides mentioned above were then measured as previously described¹⁶.

acid residues flanking each epitope, which could in turn affect both affinity and processing, or to all of these. These experiments demonstrate, however, that a potentially active epitope can be immunologically silent because of the presence of an additional epitope(s) in the protein. An otherwise immunogenic peptide can be tolerogenic when administered to adult mice in the absence of added adjuvant (M.T.S., manuscript in preparation)¹². So we tested whether administration of the joined peptide in the absence of adjuvant (that is, in tolerogenic form) would induce tolerance for both epitopes or, as is the case for induction of immunity, whether only one epitope is active. Administration of the joined peptide in tolerogenic form could induce tolerance to p12-26 (Fig. 3A, h), whereas such treatment had no effect on the immune response to pOVAD (Fig. 3A, k). By contrast, the joined peptide containing the original Val in position 327 of the ovalbumin peptide (that is, a more immunologically active epitope) was able to induce tolerance to the pOVA epitope (Fig. 3B, c), whereas the p12-26 epitope, when linked to this more active epitope, was only partially active for inducing either tolerance (Fig. 3B, b) and immunity (Fig. 2e, f). As a control, we showed that the joined peptides, p12-26 and pOVAD, were each able to induce tolerance to themselves (Fig. 3A, d, f, j and B, a), whereas pOVAD did not induce tolerance to p12-26, and p12-26 did not induce tolerance to pOVAD.

The epitope contained within p12-26 is mostly responsible for the immune activity of the joined peptide p12-26-Gly-Pro-Gly-pOVAD. So, as expected, administration of p12-26 in a tolerogenic form induced tolerance to this joined peptide (Fig. 3A, b). The immune response to the joined peptide after prior tolerization with p12-26 was strongly reduced, and the residual activity can probably be attributed to the recognition of a newly created junctional epitope(s) within the joined peptide, because the epitope contained in pOVAD remained inactive.

The finding that the joined peptide is not able to induce tolerance to pOVAD could be relevant to the mechanisms involved in autoimmunity. The presence within a self protein of non-tolerized epitopes could lead to autoimmunity when infections with pathogens containing these epitopes, or gross tissue damage that generates protein fragments, results in exposure to peptides containing such non-tolerized self epitopes in an immunogenic form.

Taken together, the results presented here indicate that the same hierarchy between epitopes exists for both immunity and tolerance. Thus, on the basis of the *in vivo* and *in vitro* experiments that we have described, we propose that such hierarchy could be regulated by the relative affinity of epitopes for the MHC-encoded class II molecule. The mechanisms that normally process and present epitopes for immunity also seem to be necessary for induction of tolerance. The presence or absence of co-stimulatory signal(s), presumably provided by the APC, might dictate the outcome of the T-cell response^{13–15}.

Received 10 July; accepted 16 November 1989.

- Guillet, J. G. *et al. Science* **235**, 865–870 (1987).
- Roy, S., Scherer, M. T., Briner, T. J., Smith, J. A. & Geffer, M. L. *Science* **244**, 572–575 (1989).
- Allen, P. M., Matsueda, G. R., Haber, E. & Unanue, E. R. *J. Immunol.* **135**, 368–373 (1985).
- Kuisaki, S., Atassi, H. & Atassi, M. *Eur. J. Immunol.* **16**, 236–240 (1986).
- Townsend, A., Goth, F. & Davey, J. *Cell* **42**, 457–462 (1985).
- Shimonkewitz, R., Kappler, J., Marrack, P. & Grey, H. M. *J. Immunol.* **133**, 2067–2074 (1984).
- Schaeffer, E. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 4649–4653 (1989).
- Sette, A. *et al. Nature* **328**, 395–399 (1987).
- Shastri, N., Miller, A. & Sercarz, E. E. *J. Immunol.* **136**, 371–376 (1986).
- Guillet, J. G., Lai, M. Z., Briner, T. J., Smith, J. A. & Geffer, M. L. *Nature* **324**, 260–262 (1986).
- Adorini, L. *et al. Nature* **334**, 623–625 (1988).
- Gammon, G. *et al. Nature* **319**, 413–415 (1986).
- Quill, H. & Schwartz, R. H. *J. Immunol.* **138**, 3704–3712 (1987).
- Lamb, J. R. *et al. J. exp. Med.* **157**, 1434–1447 (1983).
- Markmann, J. *et al. Nature* **336**, 476–479 (1988).
- Lai, M. Z. *et al. J. Immunol.* **139**, 3973–3980 (1987).
- Sauer, R. T. & Andregg, R. *Biochemistry* **17**, 1092–1100 (1978).
- Merrifield, R. B. *J. Am. chem. Soc.* **85**, 2149–2154 (1963).

ACKNOWLEDGEMENTS. This work was supported by the NIH and the ACS (M.L.G.), and by Koehst Aktiengesellschaft, FRG (J.A.S.). M.T.S. was supported by an NIH Training Grant. B.M.C.C. was supported by postdoctoral fellowships from the Cancer Research Institute and the MRC of Canada.

Reverse self-splicing of group II intron RNAs *in vitro*

Susanne Augustin, Manfred W. Müller & Rudolf J. Schweyen

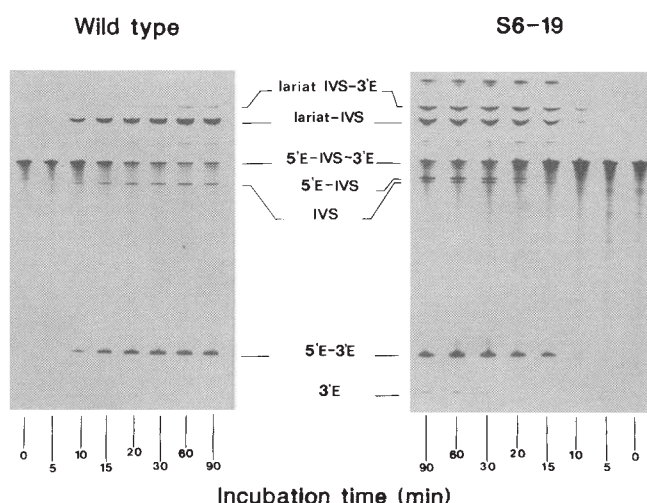
Institut für Mikrobiologie und Genetik, Universität Wien, Althanstrasse 14, A-1090 Wien, Austria

GROUP II introns, which are classed together on the basis of a conserved secondary structure¹, are found in organellar genes of lower eukaryotes and plants. Like introns in nuclear pre-messenger RNA, they are excised by a two-step splicing reaction to generate branched circular RNAs, the so-called lariats². A remarkable feature of group II introns is their self-splicing activity *in vitro*^{3–6}. In the absence of a nucleotide cofactor, the intron RNAs catalyse two successive transesterification reactions which lead to autocatalytic excision of the lariat IVS from pre-mRNA and concomitantly to exon ligation. By virtue of its ability to specifically bind the 5' exon⁷, the intron can also catalyse such reactions on exogenous RNA substrates⁸. This sequence-specific attachment could enable group II introns to integrate into unrelated RNAs by reverse splicing, in a process similar to that described for the self-splicing *Tetrahymena* group I intron⁹. Here we report that group II lariat IVS can indeed reintegrate itself into an RNA composed of the ligated exons *in vitro*. This occurs by a process of self-splicing that completely reverses both transesterification steps of the forward reaction: it involves a transposition of the 2'–5' phosphodiester bond of the lariat RNA into the 3'–5' bond of the reconstituted 5' splice junction.

FIG. 1 Comparison of products generated by *in vitro* self-splicing of wild-type and mutant S6-19 pre-mRNA. Gel-purified pre-mRNA (lane 0) of wild-type (left) or mutant S6-19 (right) was incubated under standard self-splicing conditions for different times. 5'E-IVS-3'E: full-length pre-mRNA; lariat IVS-3'E: lariat-3' exon; lariat-IVS: intron lariat; 5'E-IVS: linear 5' exon-intron RNA; IVS: broken lariat (wild type), or broken lariat and linear intron, (S6-19); 5'E-3'E: ligated exons; 3'E: free 3' exons.

METHODS. Full-length pre-mRNAs (5'E-IVS-3'E) were synthesized *in vitro* using T3 polymerase in the presence of [α - 35 S]UTP from *Eco*RI-digested plasmids BS/bl1 and BS/bl1.S6-19 as described⁶. Pre-mRNAs were gel-purified and incubated at 0.04 μ M in 40 mM Tris-SO₄, 60 mM MgCl₂, 2 mM spermidine, 500 mM (NH₄)₂SO₄ at pH 7.5 and 45 °C for self-splicing. Samples of 1.5 μ l were taken at times from 5 to 90 min as indicated and mixed with 4 μ l loading buffer (98% formamide, 10 mM EDTA, pH 7.8, 0.1% xylene cyanol, 0.1% bromophenol blue). The samples were heated for 2 min at 65 °C, snap-cooled on ice and analysed on a 5% denaturing polyacrylamide-urea gel (8 M urea; cross linking ratio, 1:30). BS/bl1 was constructed by recloning the *Sal*I-*Eco*RI insert of the initial plasmid pSP6/bl1S+ (ref. 5) into the BlueScript KS+ vector and digesting with exonuclease III to introduce a partial 5' exon deletion. Transcription of *Eco*RI-terminated BS/bl1 template gives rise to wild-type pre-mRNA comprising 24 nucleotides of the 5' exon bE1 preceded by 11 nucleotides of the plasmid, the entire intron bl1 (765 nucleotides) and the 3' exon bE2 (14 nucleotides), plus 226 nucleotides of intron bl2. BS/bl1.S6-19 was constructed by oligonucleotide-directed mutagenesis using the 21-mer 5'CCTATCACTTTTACATTG with the *Bgl*II-*Eco*RI fragment of BS/bl1 inserted into M13mp9rev (ref. 6); subsequently the mutant *Bgl*II-*Eco*RI fragment was substituted for the wild-type counterpart in BS/bl1 and analysed by DNA sequencing.

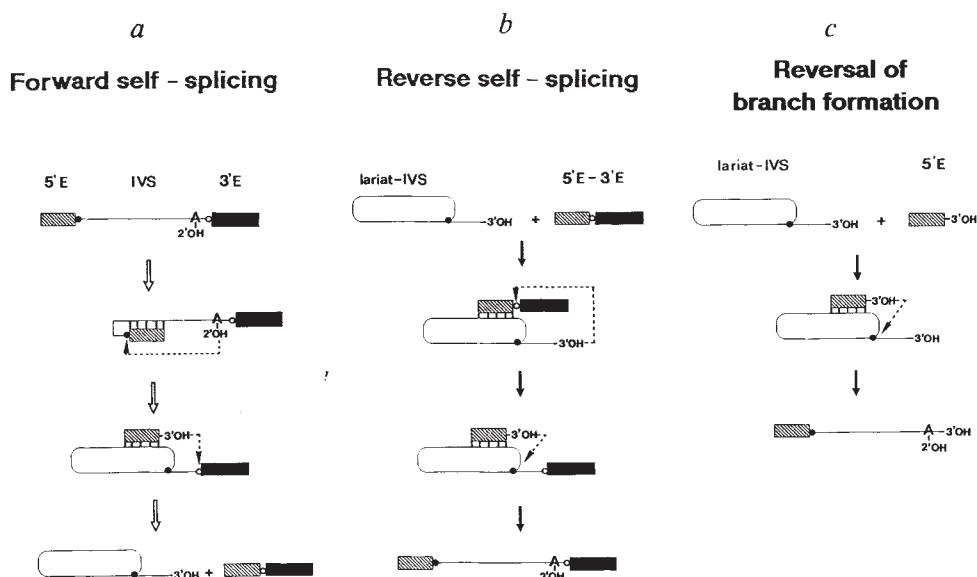
Time course experiments displayed in Fig. 1 (left) show self-splicing products of the group II intron bl1 (wild type), which originates from the yeast mitochondria apocytochrome *b* gene. The pattern of splicing products is consistent with a two-step transesterification pathway⁴ (Fig. 2a). Lariat IVS-3' exon (IVS-3'E) and 5'-exon intermediate RNAs, generated by the first phosphoester transfer, are present as minor products, and the lariat IVS and ligated exons, liberated in the second transesterification, are major products. In parallel, we analysed the mutant S6-19, in which the wild-type 3' splice site AAU:UG motif is substituted by a run of U residues. When compared with the wild-type RNA, self-splicing of mutant S6-19 pre-mRNA gives rise to different products and to changes in the



molecular ratio of some standard products (Fig. 1, right). The appearance of linear 5'E-IVS RNA and of free exons (5'E and 3'E) is novel; they accumulate during the time course (not shown for 5'E). Free exons may result from intron-catalysed reopening of ligated exons, an activity found for the wild-type IVS in incubation buffers containing high concentrations of KCl¹⁰. Exon reopening may be regarded as a partial reversal of the splicing reaction¹¹. The concentration of the lariat IVS-3'E splicing intermediate is greatly increased, being as abundant as the lariat IVS. Finally, a relatively large part of the S6-19 pre-mRNA is retained even after prolonged incubation.

The splicing pattern of mutant S6-19 RNA can be interpreted as containing products of the forward splicing pathway (com-

FIG. 2 Forward and reverse self-splicing of group II intron RNAs. Hatched and solid boxes represent 5' and 3' exons respectively; the IVS is shown as a thin line; open and filled circles mark phosphodiester bonds involved in transesterification reactions. Base pairing of exonic IBS sequences with intronic EBS sequences is indicated. *a*, Forward self-splicing catalysed by the RNA structures proceeds by a two-step transesterification pathway⁴. In the first step, the 2'-hydroxyl group of the branch adenosine attacks the 3'-5' phosphodiester bond at the 5' splice site (●) leading to 2'-5' branch formation by transesterification; the kinetic intermediate in splicing consists of two RNAs, the lariat IVS 3' exon and the 5' exon. In the second reaction step, the 3'-hydroxyl of the 5' exon exerts a nucleophilic attack at the 3'-5' phosphodiester bond at the 3' splice site (○), resulting in exon ligation and release of the lariat IVS by transesterification. *b*, Reverse self-splicing. Reintegration of the lariat IVS into the 3'-5' phosphodiester junction of the ligated exons (○) is presumed to proceed by reversal of the two transesterification steps in forward splicing. In the first step, the 3' hydroxyl of the lariat IVS makes a nucleophilic attack at the 3'-5' phosphodiester bond between the ligated exons; transesterification leads to reconstitution of the kinetic intermediate in splicing, the lariat IVS-3'E and the 5' exon. Reversal of branch formation is achieved in



the second reaction step; the 3' hydroxyl of the 5' exon attacks the 2'-5' phosphodiester bond of the lariat IVS-3'E (●), resulting in reconstitution of pre-mRNA (5'E-IVS-3'E) by transesterification. *c*, Formation of a 5'E-IVS RNA by reverse splicing. This involves the second step of the reverse splicing only, where the 5' exon exerts a nucleophilic attack at the 2'-5' phosphodiester bond of the lariat, resulting in the reconstitution of the 5' splice junction.

pare Fig. 2a), as well as RNA species that could be generated by the reverse splicing process (compare Fig. 2b, c). The abundant lariat IVS-3'E RNA may accumulate as an intermediate of either pathway. By contrast, it is difficult to reconcile the 5'E-IVS RNA with the forward splicing pathway and it has not been detected in splicing assays of wild-type RNA; this product may arise only from the reversal of branch formation, as shown in Fig. 2c. The reintegration of a group II intron into its cognate exon RNA, as formulated in Fig. 2b, is a true reversal of the two transesterification steps of the forward splicing and therefore does not involve any nucleotide cofactors. The ligated exons, provided *in trans*, are positioned at the lariat IVS by the interaction of the exon binding sites of the introns with the complementary intron binding sites of the 5' exon. Reverse splicing should be initiated by nucleophilic attack of the 3'OH group of the lariat IVS at the 5'E-3'E junction, leading to the formation of the lariat IVS-3'E RNA. Subsequently, the 3'OH group of the liberated 5' exon might make a nucleophilic attack at the 2'-5' phosphodiester bond; as a result, the 3'-5' phosphodiester bond of an authentic 5' splice site is reconstituted by transesterification. The second step, reversal of branch formation, may occur independently of the first if a free 5' exon is provided (Fig. 2c).

To test the hypothesis of reverse splicing, we purified the lariat IVS-3' exon intermediate from a mutant S6-19 splicing assay and incubated it under standard splicing conditions with an excess of free 5' exon substrate RNA. As shown in Fig. 3a, a product co-migrating with pre-mRNA (5'E-IVS-3'E) is reconstituted in the splicing assay. Wild-type IVS-3'E behaves similarly, although reconstitution is much slower (data not shown). The generation of pre-mRNA suggests that the 5' exon becomes covalently ligated to the lariat IVS-3'E in a reversal of the first transesterification step used in forward splicing, namely a reversal of lariat formation. To confirm this result, before incubation we 5' end-labelled the 5' exon substrate RNA with ^{32}P in addition to the ^{35}S label already present (Fig. 3b). This label is found exclusively in the reconstituted pre-mRNA (5'E-IVS-3'E) and with the ligated exons (5'E-3'E). This result is strong evidence that the 5' exon actually becomes ligated to the

lariat IVS-3'E. The simultaneous occurrence of the reconstituted pre-mRNA and of the ligated exons indicates that the 3' hydroxyl of the 5' exon has two potential targets for nucleophilic attack at the lariat IVS-3'E: the 2'-5' branch site and the 3' splice site (Fig. 2b and a, respectively). In wild-type RNA, the 3' splice site is apparently the preferred target, ensuring that the forward splicing reaction is highly favoured over the reverse. In mutant S6-19, the substitution of the 3' splice site by a run of U residues may hinder nucleophilic attack at this target site, but not at the branch site, so rendering the forward and reverse splicing reactions about equally efficient.

We then incubated the 5' exon substrate with free lariat IVS (instead of the lariat IVS-3'E intermediate RNA). This also gave rise to a ligation product (Fig. 3c). It co-migrates with the 5'E-IVS RNA identified as a major product in the S6-19 splicing reactions shown in Fig. 1 (right). As the 5'E-IVS product cannot result from a transesterification reaction during forward splicing and is not detected with wild-type RNA (Fig. 1, left), this product is unique to the reverse splicing reaction. The 5' exon substrate can thus react with the lariat IVS, as well as with the lariat IVS-3'E intermediate RNA; this indicates that the reversal of branch formation and subsequent ligation of the 5' exon to the IVS by transesterification does not depend on the presence of the 3' exon.

Complete reverse self-splicing was obtained by incubation of the end products of the forward reaction, that is, the excised lariat IVS of mutant S6-19 and the ligated exons (5'E-3'E). Major reaction products of this splicing assay comigrate with the lariat IVS-3'E and the pre-mRNA (5'E-IVS-3'E), indicating that both the 5'E and the 3'E are ligated to the IVS RNA by reverse splicing (Fig. 4a). When wild-type lariat IVS is used instead of mutant S6-19 IVS, similar products are obtained, but much less efficiently.

To confirm this result, ligated exons were specifically ^{32}P -end-labelled either at the 5'- or at the 3' terminus before incubation with unlabelled lariat IVS RNA. In both cases the product co-migrating with the 5'E-IVS-3'E pre-mRNA carried the ^{32}P end-label. Labelled products corresponding in size to the 5'E-IVS or the lariat-IVS-3'E were also observed (data not shown).

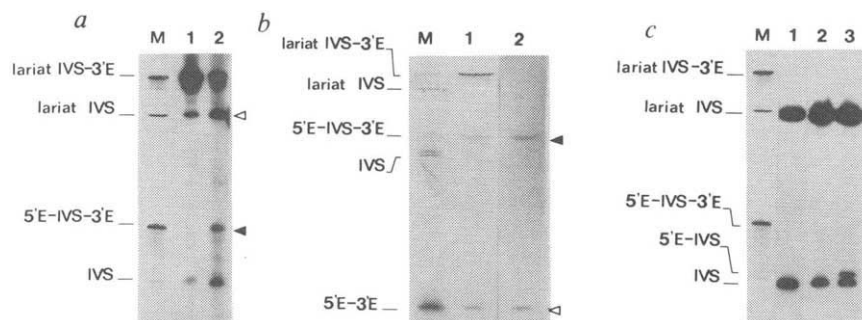


FIG. 3 Reconstitution of the 5' splice junction by reverse splicing *in vitro*. Gel-purified lariat IVS-3' exon (1 μM) or lariat IVS (1 μM), generated in splicing assays of mutant S6-19 pre-mRNA, were incubated under standard splicing conditions for 90 min with 1 μM free-5' exon substrate RNA. Reaction products were electrophoresed on denaturing 5% polyacrylamide gels. M: Self-splicing products, generated by incubating [α - ^{35}S]UTP-labelled pre-mRNA of mutant S6-19 for 15 min (a and c) or 2 h (b). Open arrows, products of the forward reaction; filled arrows, putative reverse splicing products. a, Incubation of [α - ^{32}P]UTP-labelled lariat IVS-3'E from mutant S6-19 with 5' exon substrate RNA (lane 2). 5'E-IVS-3'E, reconstituted pre-mRNA generated by the reversal of branch formation; lariat IVS, excised intron lariat generated in the second step of forward splicing. Lane 1 shows the lariat IVS-3'E before incubation; the lariat IVS and the broken lariat (IVS) were present as contaminants. b, Incubation of [α - ^{35}S]UTP-labelled lariat IVS-3'E from mutant S6-19 with ^{32}P -5' end-labelled 5' exon substrate (lanes 1 and 2). The exon-specific ^{32}P -5' end-label is seen with products co-migrating with the

5'E-IVS-3'E and with the ligated exons (5'E-3'E). We presume that these RNAs are generated by the nucleophilic attack of the 3' hydroxyl of the 5' exon at two potential targets, the 2'-5' phosphodiester bond at the branch site (reverse splicing pathway, Fig. 2b) or, alternatively, the phosphodiester bond at the 3' splice site (forward splicing pathway, Fig. 2a). c, Incubation of [α - ^{32}P]UTP-labelled lariat IVS from mutant S6-19 with 5' exon substrate (lane 3). Lane 1: lariat IVS from mutant S6-19 before incubation; lane 2: incubation of lariat IVS in the absence of substrate RNA. In Fig. 2c we illustrate the reversal of branch formation resulting in reconstitution of the 3'-5' phosphodiester bond of an authentic 5' exon-IVS junction by transesterification.

METHODS. *In vitro* transcription, incubation under *in vitro* splicing conditions, and purification of splicing products have been described in the legend to Fig. 1 and ref. 8. 5'-exon substrate RNA was 5'-end-labelled with T4 polynucleotide kinase and [γ - ^{32}P]ATP following treatment with calf intestinal phosphatase and extraction with phenol/chloroform.

FIG. 4 Complete reconstitution of 3'- and 5' splice junctions by reverse self-splicing *in vitro*. **a**, Incubation of [α - 32 P]UTP-labelled lariar IVS of mutant S6-19 with ligated exon (5'E-3'E) substrate (lanes 2); lane 2 (incubation for 15 min) and lane 2' (incubation for 1 h) give some indication of the time course of the reverse self-splicing reaction. Arrowheads mark RNA products generated by partial (lariat IVS-3'E) and by complete (5'E-IVS-3'E) reverse self-splicing. Control lanes 1 and 1': lariat IVS before and after incubation, respectively, under splicing conditions in the absence of substrate RNA. The presumed two-step transesterification pathway, leading to reintegration of the lariat IVS into the authentic 5'E-3'E junction, is diagrammed in Fig. 2b. **b**, Methods are described in the legends to Figs 1 and 3. **b**, Primer extension analysis of products generated by reverse self-splicing. After reverse self-splicing assays (**a**), RNA bands corresponding in size to the lariat IVS-3'E and 5'E-IVS-3'E were gel-purified and analysed by reverse transcription in the presence of dideoxynucleotides⁸. Lanes are labelled according to the dideoxynucleotide used in the reactions. Lettering on each side indicates the sequence complementary to the sequence ladders, corresponding to the RNA sequence. Left, Characterization of the lariat IVS-3'E intermediate RNA. Reconstituted lariat IVS-3'E was used as template in primer extension reactions with a 32 P-5' end-labelled 3' exon primer. The arrow on the left indicates the accurately reconstituted 3' splice site (3'ss); arrowheads refer to the complementary DNA termination signal of reverse transcription and to the corresponding position in the b11 sequence. Note that the reverse transcription stop, exactly one nucleotide 3' to the branch adenosine (marked by an asterisk), is a hallmark of the lariat IVS-3'E intermediate RNA (ref. 6). The branch adenosine is located 8 nucleotides 5' to the 3' splice site. Right, Characterization of the reconstituted 5'E-IVS-3'E RNA. 5'E-IVS-3'E of mutant S6-19, generated by reverse splicing was subjected to primer extension analysis with a 3' exon primer to determine the 3' splice site (left) and an intron primer to examine the 5' splice site (right). The arrows indicate the accurately regenerated 3' splice junction (3'ss) and 5' splice junction (5'ss).

METHODS. Reaction products of reverse self-splicing assays were gel-extracted and co-precipitated with 5' end-labelled intron primer or exon primer and analysed by primer extension in the presence of dideoxynucleotides⁸. The reaction products were separated on a denaturing 7.5% polyacrylamide gel. 5' End-labelling of the 3' exon primer (5'GGCGATATTATATTACC, position 814-831 in S6-19 pre-mRNA) and the intron primer (5'GGTATGATATTATAATATAC, position 51-73 in S6-19 pre-mRNA) was with [γ - 32 P]ATP and T4 polynucleotide kinase as previously described⁸.

This indicates that both exons are ligated to the IVS RNA in the reverse reaction pathway. The efficiency of the reverse splicing by S6-19 IVS is low: fully reconstituted pre-mRNA and the lariat IVS-3'E intermediate RNA were found to contain about 0.5% and 2%, respectively, of the input 32 P-3' end-label. Finally, the 5'- and 3' splice junctions of the reconstituted pre-mRNA were sequenced by primer extension analysis. The results in Fig. 4b show that both the 3'- and the 5' splice site were accurately reconstituted. It should be mentioned, however, that all splicing was under standard self-splicing conditions (compare Fig. 1); these conditions were optimized for the forward splicing reaction of wild-type RNA, but they may be far from optimal for the reverse reaction.

Our results show that group II lariat IVS RNA can integrate into a substrate RNA at a compatible location, that is, at the authentic 5' exon-3' exon junction, by the complete reversal of the forward splicing reaction. The reverse self-splicing generates the same intermediates as the forward splicing process, namely the lariat IVS-3'E and free 5' exon. As no high-energy cofactor is consumed in reverse self-splicing, it is reasonable to assume that reinsertion of the lariat IVS into ligated exons occurs by reversal of the transesterification reactions involved in forward splicing (Fig. 2a and b respectively). Group II introns, as well as group I introns, have no conserved sites of locations in otherwise highly conserved genes¹². This and other observations indicated that group II introns might have inserted into cognate sequences. As some group II introns contain open reading frames that might code for reverse transcriptases¹³, integration of the excised lariat IVS into an unrelated RNA by reverse splicing followed by reverse transcription has been favoured as a mechanism for group II IVS transposition^{14,15}. The results

here support this hypothesis by demonstrating that a group II lariat IVS can reintegrate into substrate RNAs *in vitro*. RNA-based insertion of a self-splicing IVS by the complete reversal of the forward reaction ensures that the location is compatible with subsequent excision of the intron from pre-mRNA⁹. So far we have demonstrated the RNA-based insertion at the authentic ligated exon junction only. We know, however, that pseudo 5' exon RNAs are accepted and processed by group II introns if they fulfill the minimal requirement of having a pseudo IBS1 motif, that is, a sequence complementary to the exon binding site of the IVS (ref. 8). We therefore anticipate that group II lariat insertion downstream of a pseudo 5' exon by reverse self-splicing could result in transposition of the IVS into a new location. □

Received 15 August; accepted 28 November 1989.

1. Michel, F. & Dujon, B. *EMBO J.* **2**, 33-38 (1983).
2. Sharp, P. A. *Science* **235**, 766-771 (1987).
3. Peebles, C. L. et al. *Cell* **44**, 213-223 (1986).
4. Van der Veen, R. et al. *Cell* **44**, 225-234 (1986).
5. Schmelzer, C. & Schweyen, R. J. *Cell* **46**, 557-565 (1986).
6. Schmelzer, C. & Müller, M. W. *Cell* **51**, 753-762 (1987).
7. Jacquier, A. & Michel, F. *Cell* **50**, 17-29 (1987).
8. Müller, M. W., Schweyen, R. J. & Schmelzer, C. *Nucleic Acids Res.* **16**, 7383-7395 (1988).
9. Woodson, S. A. & Cech, T. R. *Cell* **57**, 335-345 (1989).
10. Jarrell, K. A., Peebles, C. L., Dietrich, R. C., Romiti, S. L. & Perlman, P. S. *J. Biol. Chem.* **263**, 3432-3493 (1988).
11. Lambowitz, A. M. *Cell* **56**, 323-326 (1989).
12. Wolf, K. & Del Guidice, L. *Endocyt. C. Res.* **4**, 103-120 (1987).
13. Michel, F. & Lang, B. F. *Nature* **316**, 641-643 (1985).
14. Sharp, P. A. *Cell* **42**, 397-400 (1985).
15. Cech, T. R. *Int. Rev. Cytol.* **93**, 3-22 (1985).

ACKNOWLEDGEMENTS. We thank Dr M. Cotten (IMP-Vienna) for critical reading of the manuscript and Dr G. Schaffner (IMP-Vienna) for oligonucleotide primers. This work was supported by the Austrian Ministry of Science and Research.

Nucleosome positioning can affect the function of a *cis*-acting DNA element *in vivo*

Robert T. Simpson

Laboratory of Cellular and Developmental Biology, NIDDK,
National Institutes of Health, Bethesda, Maryland 20892, USA

POSITIONING of nucleosomes has been proposed as one mechanism whereby the activity of DNA is regulated: *cis*-acting elements located in linker DNA might be more accessible for interaction with *trans*-acting protein factors than they would be if they were directly associated with histones in nucleosome core particles. The eleven base pairs constituting the autonomously replicating sequence (ARS)¹ of the high-copy-number TRP1ARS1 plasmid of *Saccharomyces cerevisiae* are located in a linker region near the edge of a positioned nucleosome² and form an origin of replication³. Could nucleosome positioning render the ARS accessible for interaction with the proteins necessary for its function? I have tested this hypothesis by making deletions and an insertion to move the ARS into the nucleosome DNA and then examining the effects on ARS function. There is a marked decrease in copy number when the ARS is moved into the central DNA region of the nucleosome core particle, a region known to differ in structure and stability from the peripheral segments of nucleosome DNA.

The yeast ARS1 (ref. 1) has been extensively studied with regard to the elements that comprise this episomal replication origin (for reviews, see refs 4–6). In addition to the core sequence or A domain, other ARS elements include the B domain, extending some 70–120 base pairs (bp) towards the *TRP1* gene and containing a segment of bent DNA⁶, and the C domain, extending about 200 bp in the opposite direction^{7–9} (Fig. 1a). Mapping

of the chromatin structure of the TRP1ARS1 plasmid has defined a nuclease-hypersensitive region extending from the bent DNA region to the core ARS. Next to this, a stable, precisely positioned nucleosome was found over domain C (ref. 2). The rough coincidence between the location of the C domain and the position of a stable nucleosome led me to investigate whether the localization of the ARS core in an accessible region of the minichromosome is necessary for its function. I made deletions in the C domain designed to move the core sequence of the ARS into the nucleosome core particle.

Figure 1b shows a map of the parent plasmid (YpTA22) used here. Deletions began four base pairs clockwise from the ARS core sequence (at 871 map units) and were 10 to 80 bp long. Indirect end-label analysis of chromatin structure produced data such as those shown in Fig. 2a; a summary of the cutting sites in several experiments is shown in Fig. 2b. In all the mutants, cutting sites between the ARS core and the end of the *TRP1* gene are similar. The ARS core is susceptible in native plasmid chromatin and in the –30 and –50 deletions; this site is somewhat less accessible in the –60 and longer deletions. The cutting site at 1,022 map units (mu), which defines the right-hand border of nucleosome I (ref. 2), is resolved as a set of three cutting sites. The data summarized in Fig. 2b indicate that a nucleosome could be located between the hypersensitive region and the centre of the nuclease cut region at 1,002–1,042 mu in each of the constructs. The deletions were indeed found to move the ARS core into the positioned nucleosome.

I then investigated the effects of this relocation of the core sequence of the ARS into the nucleosome core particle on its function. As the plasmids contained the REP3 sequence from 2 μ DNA¹⁰ and were propagated in a *cir*⁺ strain of yeast¹¹, partitioning of plasmids should be nearly equal between mother and daughter cells and the copy number of the mutant plasmids was judged to be a measure of the efficiency of ARS function. Deletions up to and including 50 bp have little effect on plasmid

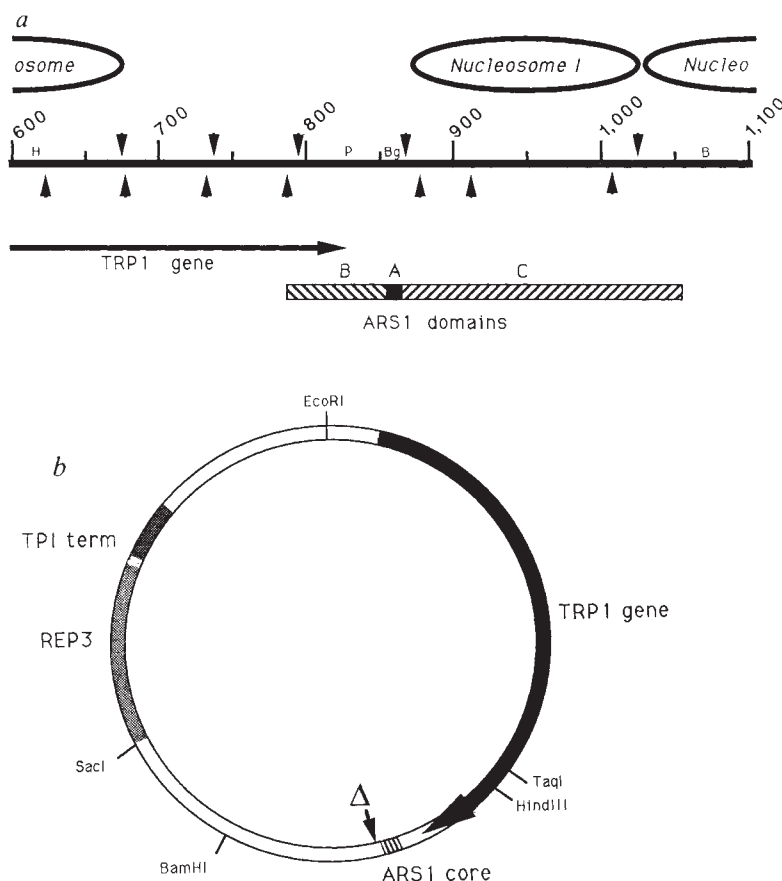


FIG. 1 Rationale and strategy for the study. *a*, Schematic representation of the organization of the ARS1 region of the TRP1ARS1 plasmid DNA and minichromosome. Locations of the core (domain A) sequences and the B and C domains in relation to the 3' end of the *TRP1* gene (the large arrowhead) are indicated. The positions of micrococcal nuclease cutting sites in naked DNA are shown by arrowheads below the centre DNA line and those in chromatin by inverted arrowheads above the line. Positions of the nucleosomes that flank the ARS hypersensitive region are indicated by open ellipses. Sites for restriction endonucleases are indicated as H, *HindIII*; P, *PstI*; Bg, *BglII*; and B, *BamHI*. *b*, Schematic diagram of pTA22, the plasmid used for deletion analysis (not to scale). The 1,453-bp TRP1ARS1 plasmid was modified by insertion of a *BamHI* site at the *NaeI* site (1,067 mu). A 100-bp DNA fragment containing the triose phosphate isomerase terminator was placed in the region of unknown function of the plasmid adjacent to a 254-bp segment of 2 μ plasmid DNA containing the REP3 element, as indicated. Deletions were at a site 4 bp clockwise from the ARS core sequence (shown as the cross-striped region near the 3' end of the *TRP1* gene), at the site indicated by the delta. The *HindIII* to *SacI* fragment containing the ARS was subcloned in pUC18 for use in mutagenesis. Deletions were made by polymerase chain reaction techniques using 24-nucleotide complementary oligonucleotides and primers for pUC regions outside the insert. The synthesized mutant DNA was cut with *HindIII* and *SacI* and ligated to a similarly digested and phosphatase-treated plasmid containing the other portion of the minichromosome. This was used directly to transform *S. cerevisiae* strain J17 (MAT α his2 ade2 trp1 met14 ura3 *cir*⁺)¹¹ to tryptophan prototrophy. Deletion lengths were confirmed by polymerase chain reaction amplification of yeast plasmid DNA sequences spanning the region of interest followed by polyacrylamide gel electrophoresis.

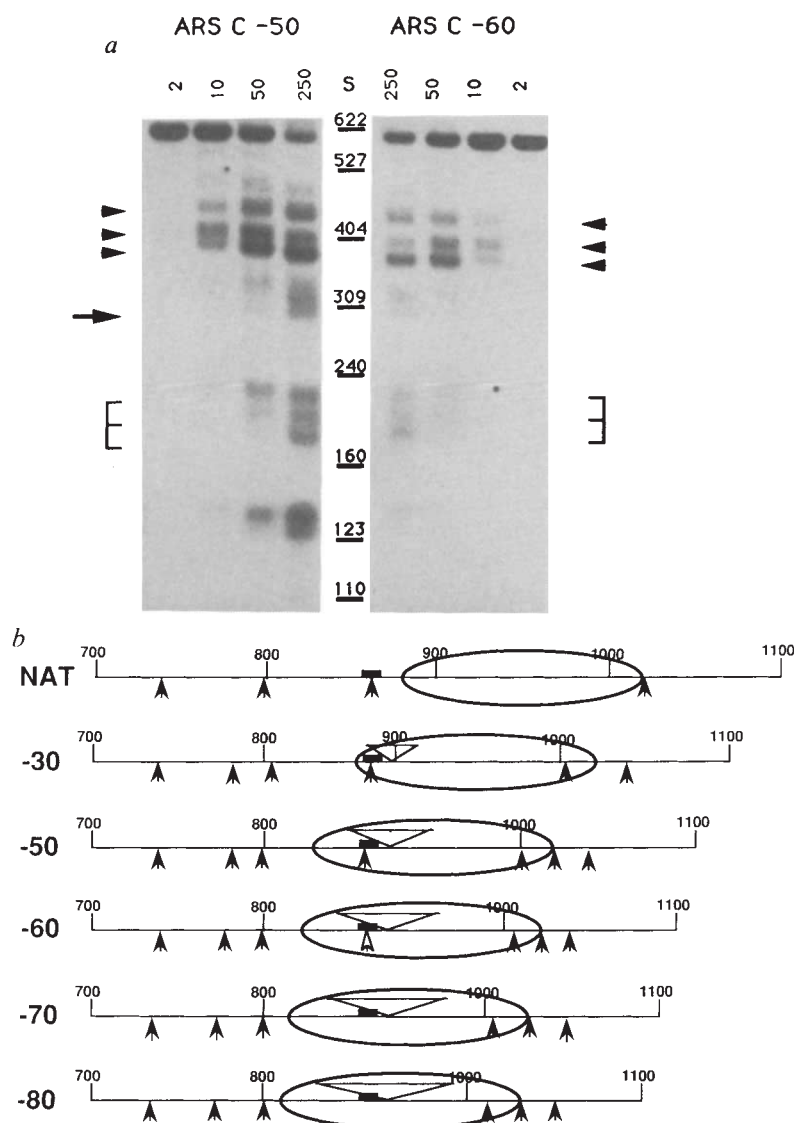


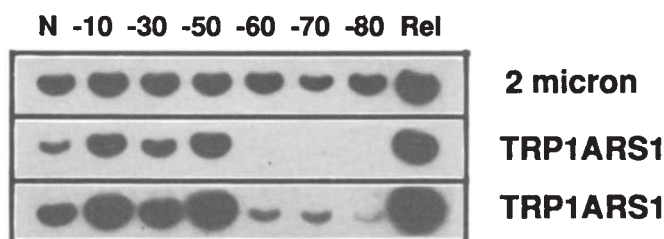
FIG. 2 Chromatin structure of the deletion mutants. *a*, Example of the indirect end-label data used to map nucleosome positions. Minichromosomes were partially purified from late log phase cultures of yeast². Samples were digested with micrococcal nuclease; DNA was purified and secondarily restricted with *TaqI* and *SacI*. For mapping chromatin structure of plasmids present in low copy number, the *BamHI*-*HindIII* fragment of the parent plasmid, which contains all of the ARS1 element, was inserted just clockwise of the TPI terminator sequences to allow high-copy-number replication and to facilitate mapping experiments. Samples shown here are from the -50 (left) and -60 (right) mutants. Four levels of micrococcal nuclease digestion are shown for each, increasing towards the centre of the panel, using the concentrations in units per ml shown above each lane. Standards were a 5'-end labelled *MspI* digest of pBR322 DNA in the centre (S) lane. The blot was hybridized with a 20-nucleotide oligonucleotide complementary to sequences near the *SacI* site. Hybridization was for 3 h at 39 °C in 1% bovine serum albumin, 1 mM EDTA, 0.5 M NaHPO₄, pH 7.2, 7% SDS²². Blots were washed 3 times at room temperature and once at 39 °C in 0.36 M NaCl, 2 mM EDTA, 20 mM NaHPO₄, pH 7.4 + 1% SDS. *b*, Schematic diagram of cutting sites and inferred chromatin structure for several of the deletion mutants. A map of nuclease cutting sites in and around the ARS hypersensitive region and nucleosome I in the parental plasmid² is shown at the top. For several of the deletions, as indicated to the left, the bottom section shows locations of nuclease cutting sites as arrowheads. These maps were aligned by assuming that the three cutting sites in the ARS hypersensitive region (left) were the same in all samples. This alignment required a change of less than 10 bp from the raw data for any sample. The location and length of deletions are shown by the inverted triangles near the centre of each DNA map. The position of the ARS core sequence is indicated by the black box for each of the deletion mutants. The inferred position of the nucleosome for each construct is shown as the open ellipse; it is primarily defined by having its right-hand border at the centre of the three nuclease cutting sites at 1,022 $\pm$ 20 map units. Note that the ARS core sequence (domain A) is moved further into the nucleosome core particle with successive deletions.

copy number, but beyond this there is a striking reduction in the copy number of the plasmid (Fig. 3). A mutant containing a transversion of the 10/11 agreement with the ARS consensus sequence at about 70 bp clockwise from the core is present at high copy number (lane labelled ARS-REL). I concluded that the deletions affected ARS function by altering chromatin structure.

A definitive test of this contention was made possible by a recent observation (S. Y. Roth, A. Dean and R.T.S., unpublished results). Studies of chromatin in yeast plasmids containing the $\alpha 2$ operator¹² in cells of different mating types indicate that the $\alpha 2$ repressor might organize chromatin structure; a positioned nucleosome is present on either side of the $\alpha 2$ binding site in α cells. I inserted a synthetic $\alpha 2$ operator in the *BamHI* site at 1,067 mu of the -60 and -80 deletions (Fig. 4a) and examined

the chromatin structure and copy number for these minichromosomes. Cutting sites in the ARS hypersensitive region and at the borders of the inserted $\alpha 2$ operator are apparent (Fig. 4b). Cutting at the former edge of the nucleosome in the -60 minichromosome (1,022 mu) is almost absent; as predicted, the $\alpha 2$ repressor has moved the nucleosome about 50 bp away from the position it occupied in the low-copy-number -60 minichromosome. Figure 4c shows the effects of these alterations of nucleosome position on copy number. Insertion of the $\alpha 2$ operator and movement of the nucleosome leads to rescue of the low-copy-number phenotype for both the -60 and -80 minichromosomes. In contrast, the $\alpha 2$ -60 plasmid exists in low copy number in α -cells where $\alpha 2$ protein is not present (data not shown). These data demonstrate that a *cis*-acting DNA element can be moved into a nucleosome and that its function

FIG. 3 Copy number analysis of ARS function in the deletion mutants. A Southern blot of plasmid DNA from the deletion mutants is shown. Yeast plasmid DNA was prepared from several of the mutant strains as well as the parent (NAT) plasmid-bearing strain as indicated. DNA was digested with *PstI*, separated by agarose gel electrophoresis, blotted to nylon and hybridized to 5' ³²P-labelled oligonucleotides specific for the TRP1ARS1 plasmid (bottom two panels) or for 2 μ plasmid DNA (upper panel). Only the relevant portion of the autoradiograph is shown for each. In the lowest panel, exposure of the TRP1ARS1 region of the autoradiograph is 10 times longer than that above; this is to show hybridization to the low-copy-number minichromosomes as an indication that the apparent reduction in copy number is not the result of a revertant in the *trp1* mutation and loss of the plasmid.



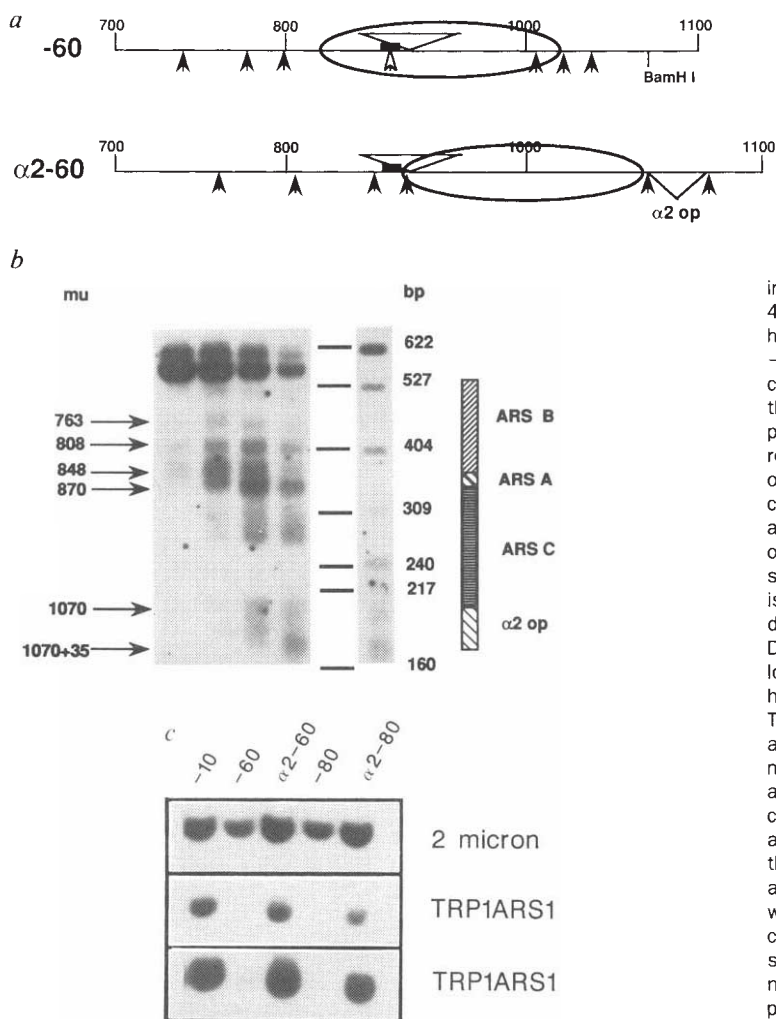


FIG. 4 Rescue of the low-copy-number phenotype of the -60 and -80 deletions by moving a nucleosome using the $\alpha 2$ operator as a chromatin structure organizer. *a*, Map of the ARS region in the minichromosomes. At the top, the positions of nuclease cutting sites for the -60 deletion, low-copy-number mutant are shown as arrowheads. The inferred location of the nucleosome in this minichromosome is indicated by the open ellipse. Below, a synthetic 36-bp DNA sequence including the $\alpha 2$ operator sequence resident 5' to the STE6 gene¹² was inserted at the BamHI site, about 50 bp away from the border of the nucleosome in the -60 deletion. The

inferred position of the nucleosome in this minichromosome (see Fig. 4b for data) is shown by the ellipse. Note that the ARS core sequence has been shifted from the central region of the core particle (in the -60 deletion) to the periphery of the nucleosome (in the $\alpha 2$ -60 construct). *b*, Indirect end-label map of the chromatin organization of the ARS region of the $\alpha 2$ -60 minichromosome. The analysis was performed as described in the legend to Fig. 2a, except secondary restriction was with *KpnI* at a site that abuts the *SacI* site. Lengths of DNA fragments are indicated to the right of the autoradiograph and cutting site positions (in map units for the parent TRP1ARS1 plasmid) are shown to the left. Four levels of micrococcal nuclease digestions of the partially purified minichromosome are shown. The basis for the smear of apparent cutting extending into the postulated nucleosome is not known. *c*, Copy number analysis of the $\alpha 2$ rescued ARS C-domain deletions. The upper third of the figure shows hybridization of a 2 μ DNA-specific probe to *PstI*-linearized total DNA from yeast, and the lower two thirds represents another portion of the Southern blot hybridized to a TRP1ARS1-specific probe. Two exposures of the TRP1ARS1 region are shown. Plasmids are displayed from left to right as the -10 deletion, the -60 deletion, the -60 deletion with a shifted nucleosome produced by introduction of the $\alpha 2$ positioning sequence, and the -80 analogues of the previous two constructs. Note the high copy number of the -10 construct, the marked reduction in the -60 and -80 deletions, and rescue of the low copy number phenotype in the $\alpha 2$ -60 and $\alpha 2$ -80 constructs by repressor-directed reorganization of chromatin structure. Exposure times for the two lower panels were 2 and 20 h, respectively. The $\alpha 2$ -60 plasmid is present in low copy number in *a*-cells, strain YNJ23 (MATa *trp1 his3 ura3 leu2 cir*⁺), showing that the introduced $\alpha 2$ operator does not supply a sequence needed in *cis* to repair ARS function in the low-copy-number phenotypes (data not shown).

is severely limited when it is located within 30–40 bp of the pseudodyad of the core particle; also, it can be moved out of this region without altering the DNA sequences formerly resident in the nucleosome, with another positioning signal, and restore function of the *cis*-acting element.

In chromatin from chicken erythrocytes and calf thymus, the 20 bp of DNA at the ends of the 146 bp in a nucleosome core particle have different properties from the central stretch of 100 bp around the core particle. These regions of DNA pull away from their normal position, apposed to the histone octamer, under thermal stress¹³. With increasing temperature, the terminal 40 bp of the core particle melt, leading to the reversible 'premelt' in chromatin denaturation¹⁴. In yeast chromatin, it is likely that this less-stable segment of the core particle is longer, the reversible premelt being larger for yeast chromatin¹⁵, and, compared with chicken core particles, there is a decrease in stability of the interactions of DNA with histones¹⁶. The crystallographic X-ray structure of the nucleosome core particle indicates that different domains exist within this basic building block of chromatin¹². Kinks or sharp bends in DNA are present at ± 15 and ± 40 bp from the pseudodyad of the core particle. The marked alteration in ARS function when this sequence is moved into the region within 30–40 bp of the core particle pseudodyad indicates that accessibility to *trans*-acting factors of the central regions of nucleosome DNA is limited. Implications of this finding for transcriptional regulation remain to be explored. Others have found partial interactions of *trans*-acting factors with nucleosome DNA *in*

vitro^{18,19} or indications of disruption of nucleosome organization when *trans*-acting factors are bound^{20,21}. This study emphasizes the need to evaluate the effects of alterations in DNA sequence (in deletion analysis of *cis*-acting elements, for example) on chromatin structure. □

Received 3 August; accepted 23 November 1989.

- Struhl, K., Stinchcomb, D. T., Scherer, S. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1035–1039 (1979).
- Thoma, F., Bergman, L. W. & Simpson, R. T. *J. molec. Biol.* **177**, 715–733 (1984).
- Brewer, B. J. & Fangman, W. L. *Cell* **51**, 463–471 (1987).
- Campbell, J. L. *A. Rev. Biochem.* **55**, 733–771 (1986).
- Campbell, J. L. *Trends Biochem. Sci.* **13**, 212–217 (1988).
- Umek, R. M., Linskens, M. H. K., Kowalski, D. & Huberman, J. L. *Biochem. biophys. Acta* **1007**, 1–14 (1989).
- Srienc, F., Bailey, J. E. & Campbell, J. L. *Molec. cell. Biol.* **5**, 1676–1684 (1985).
- Snyder, M., Buchanan, A. R. & Davis, R. W. *Nature* **324**, 87–89 (1986).
- Celniker, S. E., Sweder, K. S., Srienc, F., Bailey, J. E. & Campbell, J. L. *Molec. cell. Biol.* **4**, 2455–2466 (1984).
- Jayaram, M., Sutton, A. & Broach, J. R. *Molec. cell. Biol.* **5**, 2466–2475 (1985).
- Hill, A. & Bloom, K. *Molec. cell. Biol.* **7**, 2397–2405 (1987).
- Wilson, K. L. & Herskowitz, I. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2536–2540 (1986).
- Weisheit, W. D., Tatchell, K., van Holde, K. E. & Klump, H. *Nucleic Acids Res.* **5**, 138–160 (1978).
- Simpson, R. T. *J. Biol. Chem.* **254**, 10123–10127 (1979).
- Lee, K., Baxter, H. J., Guillemette, J. G., Lawford, H. G. & Lewis, P. N. *Can. J. Biochem.* **60**, 379–388 (1982).
- Morse, R. H., Pederson, D. S., Dean, A. & Simpson, R. T. *Nucleic Acids Res.* **15**, 10311–10330 (1987).
- Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311**, 532–537 (1984).
- Rhodes, D. *EMBO J.* **4**, 3473–3482 (1985).
- Perlmann, T. & Wrangé, O. *EMBO J.* **7**, 3073–3079 (1988).
- Richard-Foy, H. & Hager, G. L. *EMBO J.* **6**, 2321–2328 (1987).
- Almer, A., Rudolph, H., Hinnen, A. & Horz, W. *EMBO J.* **5**, 2689–2696 (1986).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).

ACKNOWLEDGEMENTS. I am indebted to colleagues in LCDB for advice and criticism, particularly S. Y. Roth, R. H. Morse and C. S. Szent-Gyorgyi. I thank Kerry Bloom for yeast strains.

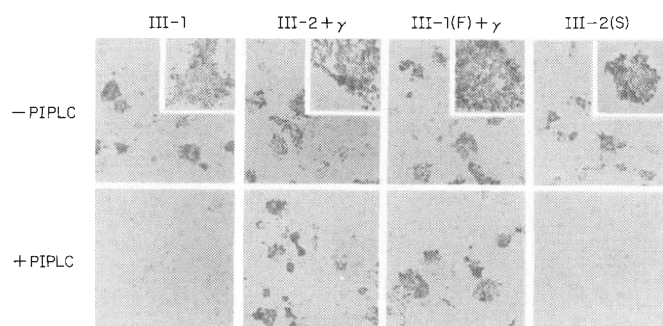
ERRATA

A single amino acid in the glycosyl phosphatidylinositol attachment domain determines the membrane topology of Fc γ RIII

Tomohiro Kurosaki & Jeffrey V. Ravetch

Nature 342, 805-807 (1989).

IN the printing process the labels of the two right-hand panels in Fig. 3 of this paper, published in the 14 December issue, were omitted. The figure is reproduced in its correct form below.

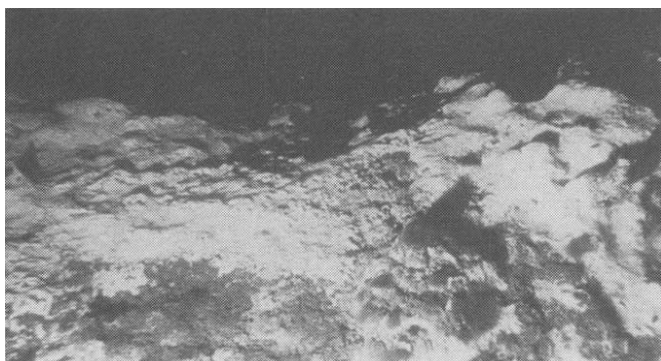


Massive natural occurrence of unusually large bacteria (*Beggiatoa* sp.) at a hydrothermal deep-sea vent site

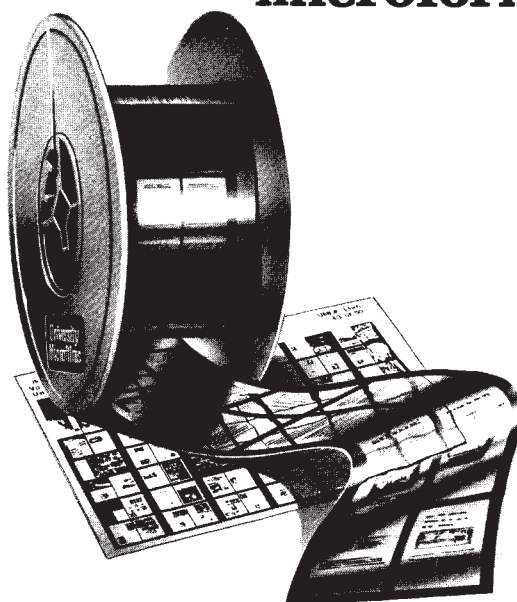
Holger W. Jannasch, Douglas C. Nelson & Carl O. Wirsen

Nature 342, 834-836 (1989).

FIGURE 1 in this letter was rotated through 180° in the published version. The correct orientation, showing *Beggiatoa* mats covering hydrothermal sediments on the sea bed, is shown below.



nature
is available in
microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International

Liquid handling in robotic workstations

A.L. Martin and J. Petracca

Addressing the limitations associated with first-generation robotics, the authors describe the emergence of second-generation machines based on an 'open architecture' with more built-in flexibility.

WHILE the 1990s hold great promise for laboratory robotics, to date, laboratory robotics have had a rather dismal track record. All too often, systems are slow, inflexible, unreliable, and expensive. The conclusion is that the most successful implementations of laboratory robots are those integrated into flexible workstations designed specifically for an application. Using simple, yet reliable, automated sample transport in a totally integrated automations environment provides cost effective, reliable and flexible automated chemistry workstations.

An important function in most sample analysis is the flexible and reliable handling of liquids. This review profiles several liquid handling products integrated under the Source for Automation (SFA) SmartCell architecture, and uses a general-purpose titrations workstation as an example of such an implementation.

The automation environment

First-generation laboratory robotics were, and still are, controlled by proprietary software running on proprietary hardware. Proprietary platforms or 'closed' architectures are deemed to be the limiting factor in the widespread use of robotics in laboratories.

Both the casual and technically adept user are limited in their ability to make a first-generation system perform. These systems work only within the boundaries of their design. Unfortunately, even in this time of increasing 'intelligence' in instrumentation, sample preparation equipment, and personal computer (PC) hardware and software, the users of first-generation equipment are still limited only to those automation modules supplied by the system manufacturer.

However, second-generation laboratory robotics are now appearing on the market as PC-based automated chemistry workstations. The shift from proprietary platforms for automation to industry-standard PC architectures is now providing laboratories with a more cost-effective, flexible and user-friendly platform for automation.

Further, first-generation systems were designed primarily as sample preparation helpers, limiting them to little more than glassware manipulators. The power of today's PC hardware and software allow second-generation workstations to be powerful data manipulators and information reporters as well. At SFA, we have implemented a PC-based open architec-

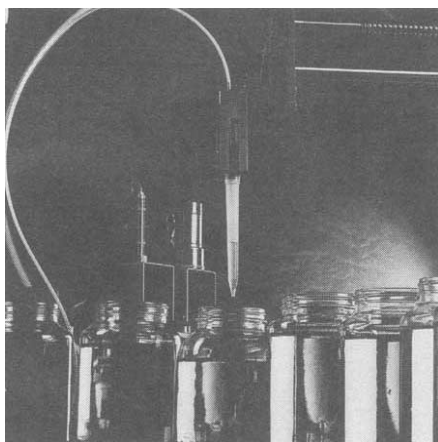


FIG. 1 Automatic transfer of a liquid sample by pipette of the fully automated robotic workstation for the BOD (Biological Oxygen Demand) assay. (Courtesy of Rohm & Haas, Bristol, Pennsylvania.)

ture written primarily in Microsoft QuickBASIC that brings a fully functional database including event-based sample handling methods to the robotic workstation.

It is the implementation of these more flexible workstations that will provide the improved reliability, repeatability and control essential for laboratories of the 1990s.

Titrations workstation

In general, the titration of a sample requires a wide variety of liquid transfer volumes, thereby requiring a wide variety of liquid transfer techniques. Various titration methodologies may call for transfers by: robotic transfer of aliquots by pipette or cannula, with or without weight verification of the transfer volume; pumped transfers using electronic balance feedback; pumped transfers using intelligent (programmable) pumps; and burette transfers.

The workstation described uses SmartCell architecture. This means that samples are logged into the sample database and scheduled into the workstations pending file. A barcode is used to tag the sample. The identifying barcode on each sample acts as a trigger for the event-based method or methods assigned to it.

Consequently, as the robot retrieves each sample for processing, its first stop is the barcode scanner. Here, the sample is identified and the appropriate 'event-based methodology' is retrieved from the database.

Event-based methods

Although the list of possible events for an automated workstation is exhaustive, we will focus on only those events directly related to liquid handling and the related products responsible for carrying out those events.

'XFER' is a completely integrated event that directs the robot to retrieve either a cannula or disposable pipette tip (depending on the event directive) and transfer a 'volume' from a 'source' to a 'destination'—the event parameters. The aliquot volume is pipetted via the SFA diluter module.

This module uses the innovative concept of a separate stand-alone power supply and communication control box to which up to eight intelligent syringe drives can be connected. The individual syringes can be placed freely about the automated workbench, thus reducing complicated plumbing and increasing flexibility in terms of bench layout.

Although the integrated event expects the SFA diluter, the event may be split into individual event components to incorporate diluters from other manufacturers. An event type 'STS', that is, a direct call to the robot, may be called to retrieve the pipetting apparatus and locate it at the source. Now, an event type of another manufacturer's intelligent syringe pump may be called to sip the aliquot. Finally, the STS would place the pipette at the destination and the manufacturer-specific dispense event would be called. Several diluters are known to fit this scheme (Cavro Scientific Instruments, Inc. Sunnyvale, California, and Hamilton Co. Reno, Nevada).

Further, in either scheme, should the 'source' or the 'destination' be a container located on an electronic balance, the difference in weight may be recorded as sample preparation audit data.

Pumped transfers are accomplished through a sequence of events similar to the former approach. However, instead of a call to a diluter event, any of the following events may be called upon.

'BFPMP', Balance Feedback PuMPing is an event accomplished by the integration of the Programmable Event Controller (PEC) with a non-intelligent peristaltic or metering pump and an electronic balance. Any number of pumps may be called on to transfer a 'volume' of liquid of known 'density' from a source to a container on an electronic balance. The PEC turns the pump on and off (actually sup-

plying or cutting off a.c. power) in intervals determined by the dynamic weight of transferred liquid approaching the target weight. Thus, virtually any pump can be used when integrated with intelligent balances (Mettler Instrument Corp. Hightstown, New Jersey, Sartorius, East Granby, Connecticut and Ohaus Corp., Florham Park, New Jersey).

'PEC' is an event carried out solely via the PEC. Non-critical volumes may be transferred from non-intelligent pumps simply via a timed switch closure at the PEC. However, intelligent burettes such as the 665 Dosimat (Brinkmann Instruments, Inc., Westbury, New York) are highly accurate and may be directly programmed for dispensing operations. The stored program may then be activated at any time via switch closure.

'PRPMP-type' events are liquid transfers carried out by intelligent pumps. The 7550 series pumps (Cole-Palmer, Chicago, Illinois) can be computer controlled via RS232C. They have command sets allowing control of pump direction, speed and number of revolutions and access to integrated calibration software.

'TITRA' events are calls to intelligent autotitrator stored methods. Once samples are prepared via the specified sequence of events, the sample may be robotically presented to the titration head where the stored titration method is carried out. As the titrator can control the complete titration process, the robot and peripherals are free to prepare the next sample while this one titrates. Autotitrators with this level of intelligence are available from Mettler Instrument Corp. and Brinkmann Instruments.

Every liquid handling technique described above is controlled through the event-based methods file. It is this methods file that allows the workstation to interchange different methods in sample preparation and analysis. This flexibility eliminates the need to batch common samples at the workstation.

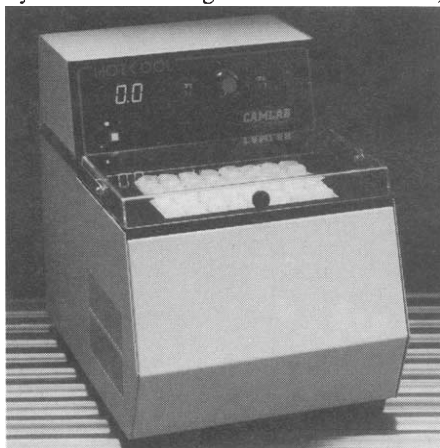
Future developments

Automated chemistry workstations of the future will bring much more sophistication to laboratory robotics, thus providing both flexible development environments and total solutions to application-specific problems. Although the titration application was chosen as an example, Smart-Cell architecture can be applied to a variety of applications. The liquid handling events described are common to many chemistries as well as pharmaceutical and biological assays, including RIA and ELISA. Other manufacturers that incorporate automated liquid handling in their products include: Beckman Instruments, Inc., Fullerton, California; Gilson Medical Electronics, Inc., Middleton, Wisconsin; Hamilton Co., Millipore Corp., Bedford, Massachusetts; Tecan, Hombrechtikon,

Liquid assets

This week's focus on instruments designed to improve liquid handling tasks range from dual-action thermomixers to complex robotic sample processors and manipulative robot arms.

THE HT-40 HotCool System from Camlab is a compact **dry block** designed to accommodate both tubes and microtitre plates (*Reader Service No. 101*). The £1,585 (UK) HT-40 incorporates two Peltier elements, which provide temperature control in the range of 0–65 °C, and eliminates, says Camlab, the need for external refrigeration or circulation systems. Measuring 200 × 300 × 250 mm,



Camlab's HT-40 HotCool dry block.

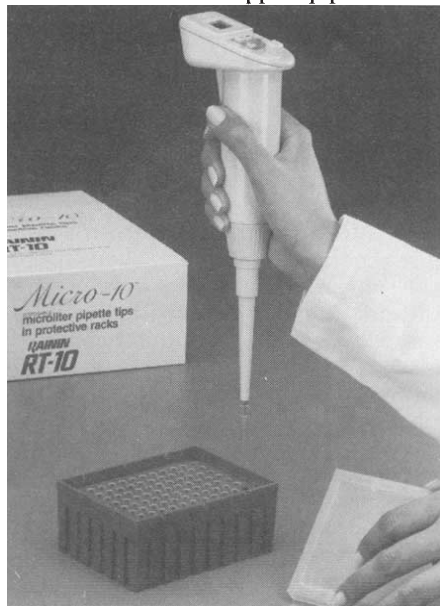
the HT-40 has a block capacity of forty 1.5-, 1.8- or 2.0-ml tubes: users can purchase an additional block for holding one 96-well microtitre plate. The dry blocks are easily removed for cleaning purposes.

The compact MC 1400 **microcentrifuge** from Hoefer Scientific is designed to perform phase separations and for spinning down precipitates, bacterial and mammalian cells, and other solids (*Reader Service No. 102*). The \$1,350 (US) MC 1400's aluminium rotor can accommodate eighteen 1.5- or 2.0-ml tubes, although 0.25-, 0.4- or 0.5-ml tubes can be used with an adapter. The MC 1400 can reach a maximum speed of 14,000 r.p.m., maximum RCF of 16,000 g in 11 seconds, says Hoefer. The rotor comes to a stop in 12 seconds — only then will the safety latch release and the lid open. A ventilation system built into the lid maintains tem-

perature at 30 °C or below — even after successive 30-minute spins. Other features include a 30-minute timer and a momentary on/off switch for manually-controlled pulses.

peratures at 30 °C or below — even after successive 30-minute spins. Other features include a 30-minute timer and a momentary on/off switch for manually-controlled pulses.

EDP2 Micro-10 is the new **battery-operated microvolume pipette** from Rainin Instrument Co. (*Reader Service No. 103*). The computer-controlled piston movement allows volumes as low as 0.5 µl to be dispensed. According to Rainin, the EDP2 Micro-10 is designed with minimum air space within the pipette shaft. The titanium-tipped pipette shaft



EDP2 Micro-10: programmable, battery-operated microvolume pipette.

provides rigidity, chemical inertness and a uniform surface for the sealing of Micro-10 tips. The pipette, costing \$245 (US), can be calibrated to dispense multiple samples with the speed, volume and mode entered via the keypad and displayed on the LCD. Other design features include a pushbutton tip ejector for the safe removal of contaminated tips, and automatic shutoff if the pipette has not been used for three minutes.

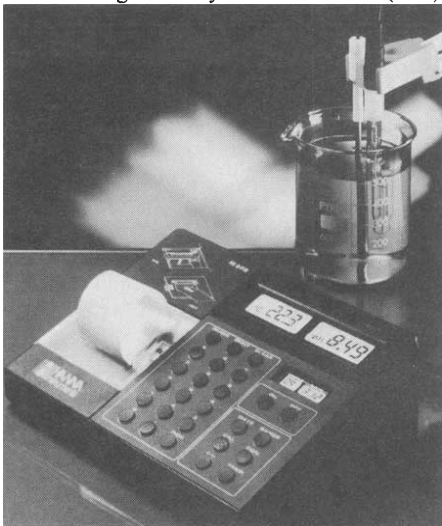
Switzerland; and Zymark Corp., Hopkinton, Massachusetts. The laboratory of the 1990s must be able to draw on the industry as a whole to realize its automation goals. The development of open-architecture automation platforms is the only way to combine state-of-the-art sample preparation equip-

ment and sample and data handling techniques into synergistic workstations. □

Arthur Martin is President of Source for Automation, Inc., and John Petracca is Leader of the Product Development Team with SFA, 115 Cedar Street N-3, Milford, Massachusetts 01757. For more information, fill in reader service number 100.

pH meters and more

Orme is distributing Hanna **bench pH meters**, which measure pH, millivolts and temperature and are supplied complete with electrodes and a full range of accessories (*Reader Service No. 104*). Two models are available: the microprocessor-controlled HI 8417 or the HI 8418 with full data management system. The £262 (UK)



Multipurpose pH meter from Orme.

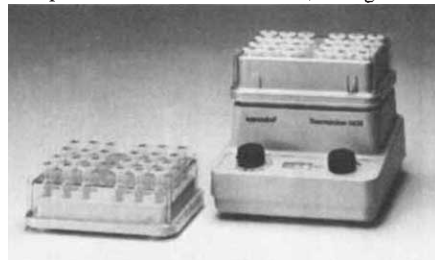
HI 8417 has a dual pH and temperature display, automatic or manual temperature compensation and instant access to the six standard buffer values in the memory. The HI 8418, costing £412 (UK), has four digital displays for temperature, pH, sample number and date, an integral printer for permanent record keeping and an audible pH alarm.

Metrohm is featuring the new all-in-one multipurpose 691 **pH meter** (*Reader Service No. 105*). Using electrodes for a variety of applications, the 691 pH meter can measure pH, voltage, with and without I_{pol} , and temperature. Automatic buffer regulation reduces the risk of incorrect calibration, says Metrohm. Calibration data can be stored in the instrument. The \$695 (US) version comes without an RS232C interface, and this cannot be retrofitted. The \$795 (US) version is supplied with the interface and can communicate with printers or PCs. A range of combined and platinum electrodes, and resistance thermometers are available, priced between \$94–305 (US).

Shaken and stirred

If mixing and incubating samples are part of your daily routine, Eppendorf recommends the 5436 and 5437 **dual-purpose thermomixers** (*Reader Service No. 106*). The 5436 and 5437 thermomixers both hold 24 microcentrifuge tubes; the 5436 takes 1.5-ml tubes and the 5437 2.0-ml tubes. Users can set the temperature in 1°C increments from a minimum of 5–10 °C above ambient to a maximum of 95 °C. The mixing intensity can be selec-

ted in the range of 500–1,400 min^{-1} . An LED displays a readout of selected nominal temperature and the difference between this and the actual block temperature. Transfer racks, designed to

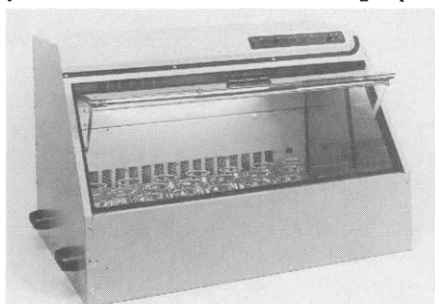


Eppendorf's dual-action thermomixer.

fit tightly in the thermomixer to maximize heat transfer, can be stacked and autoclaved.

For the safe magnetic stirring of liquids in hazardous areas, Bel-Art Products recommends the non-electric Scienceware **water or air-turbine magnetic stirrer** (*Reader Service No. 107*). The \$23-50 (US) Model F37004 stirrer and tight-fitting but separable reservoir can be operated on either air or water pressure: the minimum operating pressure is 3 p.s.i. The upper section, which can be used separately for air operation, houses a turbine with high-strength magnets for rotating a spin bar. When the stirrer is powered by water pressure, the upper section fits onto the lower reservoir section and used water collected in the reservoir is drained through a three-foot tube.

Infors of Switzerland has introduced a compact benchtop **incubator shaker** — the RFI-100 — which is suitable for high-temperature applications (*Reader Service No. 108*). The shaker unit's magnetic drive has, says Infors, a space-saving design, no wearing parts, is 'noiseless' and generates no surplus heat. Temperature, controlled by means of a proportional controller and Pt-100 sensor, can be adjusted on a continuous scale using a pot-



Infors' RFI-100 space-saving incubator/shaker keeps cultures in top condition.

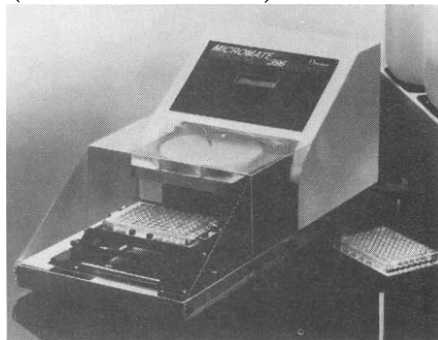
entiometer. The shaker can operate over a temperature range of 5 °C above ambient to 100 °C, without the need to use the built-in cooling system. The unit's shaking table will accept a type-F tray holding, for example, 73 100-ml Erlenmeyer flasks, or eight 2,000-ml flasks. The speed of the

shaker can be set to between 10–400 r.p.m. Other features of this front-opening design include a built-in humidifier, fluorescent lighting and corrosion-resistant components.

Microplate washers

The latest addition to Denley Instruments' Welltech range of microplate instrumentation is the fully automated **stacking microplate washer** — Wellwash 5 (*Reader Service No. 109*). Wellwash 5 can process simultaneously up to ten microplates in the 8- or 12-well format. Wash sequences are determined from plug-in cards, and variable parameters include wash volume (50–750 μl), strip or microplate order, number of washes (up to seven per plate), and soak time. Selecting the adjustable program card allows the user to vary all parameters and set up user-defined protocols. Wellwash 5 is equipped with a low-level wash-fluid sensor, which alerts the user of the need to replace reagent bottles, and a 'dispense-only' option for filling plates with volumes ranging between 20–300 μl .

The fully automated Micromate 396 **microplate washer** from Packard Instrument Co. is designed to wash 96 wells simultaneously, but can also take strips (*Reader Service No. 110*). Micromate 396



Micromate 396 — washes wells and strips.

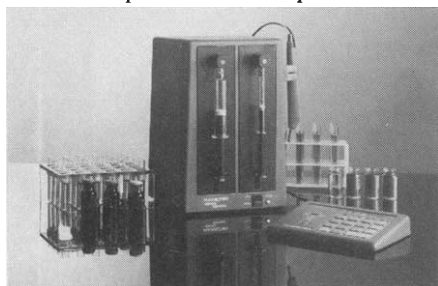
can be configured for dispense volume (100 to 999 μl), soak time (5 seconds to 10 minutes), total number of aspirate and dispense cycles (1 to 9), shaking and agitation. Costing \$3,950 (US) for the basic system, the instrument's battery-backed memory can store up to five wash programs entered via the 12-key touch panel. Additional features include position sensing to determine optimal aspiration tube depth, overflow protection to minimize carryover to the next tube, and overnight protection of aspiration and dispensing tubes to prevent clogging. A vacuum pump or standard in-line systems can be used for aspiration purposes and a pressurized fluid delivery pump or gravity design can be used for dispensing.

Liquid dispensers

Watson-Marlow is distributing Flexicon of Denmark's Flexfill CCAD80 and Flexfill 12 **autodispensing machines** (*Reader Ser-*

vice No. 111). Based on the peristaltic principle, Flexfill CCAD80 and Flexfill 12 differ only in dispensing capacity: 0.5 to 250 ml and 20 ml to 4 litres, respectively, although fill volumes can be doubled by the addition of a second dispensing pump-head. When using a single dispensing head, flow rates for the Flexfill CCAD80 can be varied between 2.5–1,500 ml min⁻¹. The design of the dispensing head ensures a filling volume that is accurate to ± 0.5 per cent, says Watson-Marlow. Operating instructions, entered via a touch-sensitive keyboard, are shown on the digital display. The autodispenser is supplied with a single dispenser head, two metres of tubing and a manual.

Hamilton has introduced the Microlab 960 **dual-syringe, dual-valve diluter/dispenser** — the latest addition to the Microlab 900 Series (*Reader Service No. 112*). Microlab 960 can be computer controlled by commands downloaded from an IBM computer or compatible. The

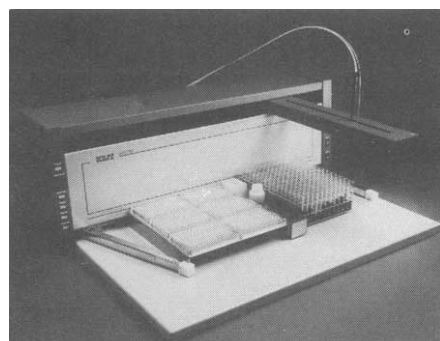


The Microlab 960 dispenser/diluter from Hamilton can be computer controlled.

instrument will then repeatedly execute the commands even when the controlling device is removed, which is useful in laboratories where changes to liquid handling protocols are infrequent. Alternatively, instrument functions can be controlled remotely with a hand probe or footswitch. The Micromate 960 can be supplied with the Microlab 900 Digital Controller, which allows the user to create and store up to 20 programs. Special features of the Microlab 960 include variable syringe drive speeds, programmable wash cycles, air gaps to minimize sample/sample and sample/reagent cross-contamination, and a volume capacity of between 1 μ l and 10 ml per syringe drive.

Lab automation

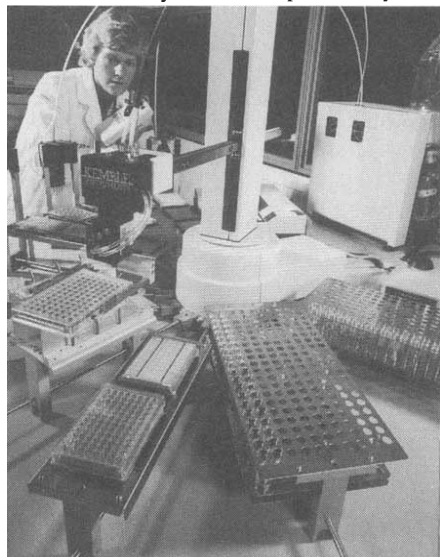
Repetitive liquid handling tasks can be performed with the minimum of user intervention using the Beeline 2 **robotic sample processor** from Hook and Tucker Instruments (*Reader Service No. 113*). Beeline 2 is available as a fully programmable PC-driven version for laboratories using a variety of assays, where the PC controls all probe movements in the x,y,z axis and/or the operation of a twin syringe diluter. Alternatively, Beeline 2 can function under its own control, storing up to nine user-defined protocols. Both



Beeline 2 robotic sample processor.

models feature a single-probe sampling tip that can be programmed to dip to a certain level, or can be fitted with an optional capacitance, conductivity or optical liquid-level sensor. Beeline 2 is also suitable for dual-probe operation. Other features include a double-chamber wash bath to keep the probe tip contamination-free and a large working area that can accommodate tubes, microtitre plates and racks.

Autolab 500 is the new **robot for automating liquid handling** from the Kemble Instrument Co. (*Reader Service No. 114*). The robot reduces contact with hazardous substances and provides increased accuracy with improved reproducibility, says Kemble. Using the robot, a range of tools can be selected without user intervention, including sampling and dispensing probes, tube and plate grippers, washers and lid handlers. The robot is controlled by a mouse-operated system



Automate lab tasks with Autolab 500.

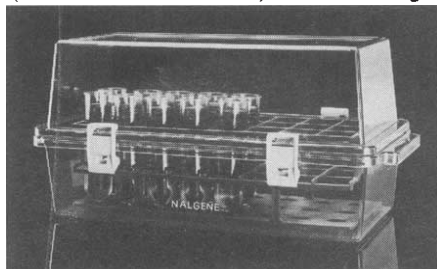
and assay routines can be stored on computer. Autolab 500 can interact with peripheral equipment such as mixers and shakers, and interface with a range of computers and barcode identification systems. Stated applications include sample preparation, blood grouping, RIAs and ELISAs.

The 232/401 **automated sample processor and injector** from Gilson Medical

Electronics combines high-capacity sample preparation and injection with column switching capabilities into one unit (*Reader Service No. 115*). The 232/401 can process up to 540 samples in a single run, providing, says Gilson, 24 hours of unattended operation. The machine features a stationary rack system that holds five separate racks, and provides access to vials in any x, y, or z pattern. An 8-KByte memory allows the user to create, store and run programs tailored to individual needs. The 232/401 can be configured to virtually any HPLC system. Using the second switching valve and pre-stored programs, the 232/401 can perform various procedures that facilitate the preparation, injection and collection of complex samples: backflushing, on-line sample cleanup, sample enrichment and concentration and collection of up to six fractions per sample.

Safety aids

Protect yourself when transporting potentially hazardous substances with the new Bio-Safe **closed-system carrier** from Nalge (*Reader Service No. 116*). Made of tough,



Nalge's Bio-Safe closed carrier provides protection from biohazards.

break-resistant polycarbonate, the Bio-Safe carrier can accommodate Nalgene Unwire test tube racks or other racks with similar dimensions. Other design features include easy-to-grasp moulded-in handles, clamps that ensure the carrier is kept closed with a leakproof seal, and clear visibility of the contents. In the event of an accidental spill, exposure to biohazards can be avoided by autoclaving and disposing of the carrier without the need to open the lid. Bio-Safe carriers cost \$36.10 (US) each.

Elkay Products has launched a new line of Safe 'N Ezee **biohazard bags** with warnings printed in English, Spanish and French (*Reader Service No. 117*). Safe 'N Ezee bags are made from high-density polyethylene material and will withstand autoclaving up to 250 °C, says Elkay. Four sizes are available: 12 × 20, 18 × 26, 24 × 30 and 40 × 45 inches.

These notes are compiled by Diane Gershon from information provided by the manufacturers. To obtain further details, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.